PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classificati n 7 : G06F 17/00	A1	(11) International Publication Number: WO 00/65472 (43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/US00/11296

(22) International Filing Date: 26 April 2000 (26.04.00)

(30) Priority Data:
60/131,105 26 April 1999 (26.04.99) US
60/175,075 7 January 2000 (07.01.00) US

(71) Applicant: SURROMED, INC. [US/US]; 1060 E. Meadow Circle, Palo Alto, CA 94303 (US).

(72) Inventors: RINGOLD, Gordon; 12004 Adobe Creek Lodge Road, Los Altos Hills, CA 94022 (US). DIETZ, Louis, J.; 550 S. Shoreline Boulevard, Mountain View, CA 94041 (US). KANTOR, Aaron, B.; 37076 Arden Street, Newark, CA 94560 (US). NATAN, Michael, J.; 765 San Antonio Road, #56, Palo Alto, CA 94303 (US). BRUNKE, Karen, J.; 11 Somerset Court, Belmont, CA 94002 (US). ALLISON, Anthony; 2513 Hastings Drive, Belmont, CA 94002 (US).

(74) Agents: SWANSON, Barry, J. et al.; Swanson & Bratschun, L.L.C., Suite 330, 1745 Shea Center Drive, Highlands Ranch, CO 80126 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: PHENOTYPE AND BIOLOGICAL MARKER IDENTIFICATION SYSTEM

(57) Abstract

A phenotyping system for obtaining multiple parameters of an organism in order to full characterize said organism. Said phenotype comprising the results of at least 20 assays relating to cell populations and/or cell associated molecules, the results of at least 20 assays relating to soluble factor and clinical parameters.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PHENOTYPE AND BIOLOGICAL MARKER IDENTIFICATION SYSTEM

SCOPE OF THE INVENTION

The present invention provides a phenotype and biological marker identification system and methods for identifying and using novel patterns of biological markers related to disease, disease progression, response to therapy and normal biological functions. The discovery and use of novel patterns of biological markers will result in more cost-effective drug development, including the improvement of patient selection in clinical trials and the identification of therapeutics with greatly improved safety and efficacy. Phenotype information and biological markers can also be used in diagnostic applications.

BACKGROUND OF THE INVENTION

As a result of recent innovations in drug discovery, including genomics, combinatorial chemistry and high throughput screening, the number of drug candidates available for clinical testing exceeds the pharmaceutical industry's development and economic capacity. In 1998, the world's top pharmaceutical and biotechnology companies spent more than $50 billion on research and development, more than one-third of which was spent directly on clinical development. As the result of a number of factors, including increased competition and pressure from managed care organizations and other payors, the pharmaceutical industry is seeking to increase the quality, including the safety and efficacy of new drugs brought to market, and to improve the efficiency of clinical development.

Recent drug discovery innovations, therefore, have contributed to a clinical trials bottleneck. The numbers of therapeutic targets being identified and lead compounds being generated far exceed the capacity of pharmaceutical companies to conduct clinical trials as they are currently performed. Further, as the industry currently estimates that the average cost of developing a new drug is approximately $500 million, it is prohibitively expensive to develop all of the potential drug candidates.

The pharmaceutical industry is being forced to seek equivalent technological improvements in drug development. Clinical trials remain very expensive and very risky, and often decision making is based on highly subjective analyses. As a result, it is often difficult to determine the patient population for whom a drug is most effective, the appropriate dose for a given drug and the potential for side effects associated with its use. Not only does this lead to more failures in clinical development, it can also lead to

approved products that may be inappropriately dosed, prescribed, or cause dangerous side effects. With an increasing number of drugs in their pipelines, pharmaceutical companies require technologies to identify objective measurements of a drug candidate's safety and efficacy profile earlier in the drug development process.

One approach to deal with the mass of information and technologies is to break away from the traditional methods of drug identification and development. As a variety of different analytical, clinical and information handling technologies continually advance, it may be possible to develop a phenotype for an individual or population that allows for an unprecedented systematic evaluation of such individual or population. The phenotype for a given individual includes, in theory, all measurable characteristics of such individual at all points in time. One use of such phenotype information is the identification of biological markers.

Biological markers are characteristics that when measured or evaluated have, inter alia, a discrete relationship or correlation as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention. Pharmacologic responses to therapeutic intervention include, but are not limited to, response to the intervention generally (e.g., efficacy), dose response to the intervention, side effect profiles of the intervention, and pharmacokinetic properties. Response may be correlated with either efficacious or adverse (e.g., toxic) changes. Biological markers include patterns of cells or molecules that change in association with a pathological process and have diagnostic and/or prognostic value. Biological markers may include levels of cell populations and their associated molecules, levels of soluble factors, levels of other molecules, genotypic information, gene expression levels, genetic mutations, and clinical parameters that can be correlated with the presence and/or progression of disease.

In contrast to such clinical endpoints as disease progression or recurrence or quality of life measures (which typically take a long time to assess), biological markers may provide a more rapid and quantitative measurement of a drug's clinical profile. Single biological markers currently used in both clinical practice and drug development include cholesterol, prostate specific antigen ("PSA"), CD4 T cells and viral RNA. Unlike the well known correlation between high cholesterol and heart disease, PSA and prostate cancer, and decreased CD4 positive T cells and viral RNA in AIDS, the biological markers correlated with most other diseases have yet to be identified. As a result, although both government agencies and pharmaceutical companies are increasingly seeking development

of biological markers for use in clinical trials, the use of biological markers in drug development has been limited to date.

Although there are many potential biological markers, there is limited technology that is capable of sorting through the vast amounts of information needed to establish the correlation of the biological markers with normal biologic processes, disease, disease progression and response to therapy. Phenotyping requires the instrumentation and assays required to measure hundreds to thousands of parameters, an informatics system to allow this data to be easily accessed, software to correlate the patterns of information with clinical data and the ability to utilize the resulting information in the drug development process. The present invention provides such a technology.

SUMMARY OF THE INVENTION

The present invention relates to phenotyping an organism or a class or subclass of organisms. The present invention also includes the identification of biological markers that are measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention. This invention includes technology capable of providing quantitative, sensitive reproducible and rapid measurements of multiple and diverse biological markers that could accurately profile an organism's phenotype or a patient's disease status and response to therapy. Further, because blood is the single most information rich tissue and is easily and readily accessible for testing, the invention focuses on identifying biological parameters from small samples of blood. The invention includes a multidisciplinary format comprising three principal elements: instrumentation, assay development and clinical informatics.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic representation of the types of information that are assimilated to obtain one embodiment of a biological marker identification system.

Figure 2 depicts a schematic representation of the improved MLSC instrument of the invention (term "SurroScan" instrumentation).

Figure 3 depicts the integrated information infrastructure for analyzing the data obtained in the present invention.

Figures 4 A - C depict the results obtained in Example 1 showing that $CD27^+$ and $CD27^-$ CD8 T cells vary among samples. Blood samples from three different donors (Figures 4A, B, and C) were stained with Cy5 anti-CD27 and Cy5.5 anti-CD8.

Figures 5A and B depict robust cellular measurements with 2-color MLSC. Figure 5A demonstrates the consistency of CD8 T cell counts from 6 different capillaries. Cy5.5 anti-CD8 was combined with a different Cy5 conjugated antibody for each of the capillaries (anti-CD3, CD25, CD7, CD45RA, CD62L, CD69). Fifty different blood samples were analyzed. The box-and-whiskers plots show that the distributions of cell counts are very similar for each of the capillaries. Pair-wise linear regression also shows a high degree of consistency for these assays (data not shown). Figure 5 shows the consistency of two measures of B cells, one with Cy5.5 anti-CD20 and one with Cy5.5 anti-CD19. The 95% confidence interval (dotted line) of the linear regression includes a slope of 1 and the fit has a correlation coefficient of 0.97.

Figure 6 shows a classification matrix comparing CD8 T cells and CD4 T cells in RA patient samples and blood bank samples.

Figures 7A and B show results of a three color cellular assay on the SurroScan instrument.

Figures 8A - C shows the results of staining intracellular molecule as measured with MLSC technology.

Figures 9A - C show the results of a 3 detection channel analysis using MLSC technology.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the phenotyping of an organism or a class or subclass of organisms. In theory, the phenotyping of an organism includes obtaining all measurable characteristics of said individual, past and present. While the complete phenotyping of any organism is not practical or even possible, the phenotyping disclosed and described herein provides an unprecedented quantity of an unprecedented number of types of parameters or characteristics so as to provide a resource of information that will allow for the analysis of normal biological functions, disease, disease progression and changes associated with virtually any perturbation to the organism.

One utility of the phenotyping system taught by the present invention is to the identification of biological markers for normal biological processes, diseases or medical

conditions. In order to perform this aspect of the present invention it is necessary to have i) biological information from a population of individuals, ii) an adequate amount of data from each individual, preferably obtained by multiple sampling over time, and iii) an information storage and retrieval system that a) can integratively incorporate a wide variety of types of information and b) can perform meaningful correlation analysis of the disparate types of data. Figure 1 depicts information that is useful to create a biological marker identification system.

Disease and disease progression involves the complex interplay of both genetic and environmental factors. The present invention has the potential to identify and trace changes in patterns of biological markers reflecting both genetic and environmental factors from small samples of blood. Furthermore, the present invention helps decipher genetic components of disease susceptibility, disease progression and response to therapy.

The present invention is capable of monitoring cells, proteins, organic molecules, genotype, soluble factors, clinical and environmental factors, all of which have been used as biological markers in drug development and as disease markers. Examples of known biological marker include the monitoring for decreases in CD4 positive T cells and viral RNA levels in AIDS, elevated cholesterol levels as an accepted biological marker for heart disease and changing levels of PSA as a protein marker found in the blood of prostate cancer patients.

Since the biological characteristics or parameters that might be discovered to be a biological marker or part of a marker "grouping" are often not predictable, it is essential that the appropriate database contain information regarding as many parameters as possible.

The present invention extends to a phenotype of a given organism, methods for assembling such phenotype and methods for utilizing such phenotype. The phenotype of an organism or class or subclass or organisms comprises a large compilation of data relating to the organism or class or subclass of organisms. The novel aspect of the present invention lies in the disparate nature of the data and the quantity of data from each of the various categories of data available on an organism. A phenotype can only reach its full usefulness if the data defining the phenotype is extensive. For example, a phenotype for a human patient containing a standard blood profile and clinical factors routinely obtained from a physical examination can not provide enough information to fully exploit such a phenotype. Although the assays involved and data obtained are within the scientific and

clinical capabilities of the art, obtaining all of the information from a single organism is a novel task. Although the handling and maintenance of a phenotype lends itself to computerization, a given phenotype can be kept in traditional formats. Manipulating phenotypes to identify a biological marker or to observe the effect of a perturbation in the organism is, of course, greatly simplified by the use of computational analysis via a computer. As described above, the complete phenotyping of an organism would include literally thousands or possibly millions of data points. In the preferred aspects of this invention, a phenotype comprises greater than 40 biological parameters, more preferably greater than 100 parameters, and most preferably greater than 200 different parameters, and in some cases greater than 300 different parameters. The phenotype must contain biological parameters that include information from cellular assays, soluble factor assays and clinical information. In the preferred embodiment, the results of at least 20 cellular assays incorporating measurements of at least 20 cell populations and/or cell associated molecules and the results of at least 20 soluble factor assays are included in the phenotype, along with clinical information. In more preferred embodiments, the results of at least 40 cellular assays incorporating measurements of at least 40 cell populations and/or cell associated molecules and at least 40 soluble factor assays are included, preferably with an extensive battery of clinical and environmental parameters. In preferred embodiments of the invention there are included more than 20 clinical parameters, preferably more than 40 and in some cases more than 60 clinical parameters.

A rich and readily accessible source of biological information for a patient is the blood. At the present time, there are over 200 identified discrete leukocyte cell surface antigens with identified antibodies. In addition, there are literally thousands of proteins and other soluble factors and small molecules that can be identified in blood. The problem, therefore, is not in finding enough informational content in the blood, but in efficiently extracting all of the available information from limited quantities of blood.

Many levels of biological markers may vary widely from individual to individual. In many cases, such variations may be random, but this may not always be the case. For example, in some situations baseline levels may be individual specific, and only by taking multiple readings from an individual would it be possible to identify a biological marker. Although it may not be likely that a baseline would be established for a healthy individual, there may be valuable information gained from the variations over time in a given individual that has a disease or medical condition. For example, a patient with rheumatoid

arthritis may show interesting variations when off or on medicine, or when exhibiting a severe flare-up of symptoms. If such longitudinal correlations exist, review of the longitudinal data of other similarly situated patients could confirm valuable biological markers associated with the disease. When longitudinal data over an extended period of time exists, the number of individuals necessary for the analysis to be statistically significant can be relatively small.

An additional application of the present invention is in monitoring dose response studies. In this embodiment, a population of individuals is evaluated before and after the administration of drug and after increasing doses of the drug. In this embodiment, the selected population may be healthy individuals, and the anticipated biological dose response endpoint is toxicity or side effect profiles. In embodiments where the individuals have a particular disease or medical condition, markers may be identified for efficacy along with the negative effects of the drug. By evaluating the information from individuals before and after administration of drugs it will be possible to identify markers or marker groupings associated with administration and response to the drug. In some situations, such markers could be used as an endpoint for clinical studies. For example, in contrast to such clinical endpoints as disease progression or recurrence or quality of life measures (which typically take a long time to assess), biological markers may provide a more rapid and quantitative measurement of a drug's clinical profile.

In other embodiments of the present invention, longitudinal studies of individuals receiving a drug or treatment for the prevention or treatment of a disease or medical condition could constitute the population of individuals being evaluated. By correlating biological indicators of individuals before they receive treatment with subsequent clinical observations, it will be possible to identify biological markers associated with those members of a potential patient population that will most benefit from the treatment therapy. In such a manner, expensive treatments can be limited to the subpopulation of patients most likely to benefit from the treatment.

Another application of the present invention is the use of biological markers to identify patients who have very early clinical signs of a disease. This would be extremely valuable for a multitude of disease states where a patient may have "subclinical" signs and symptoms which are not severe enough to bring them to a doctor's office. However, if a patient had a marker which was discovered in their blood and they were advised to seek medical attention, their "subclinical" signs could be identified as their earliest phenotypic

presentation of a disease. For many diseases, it is extremely advantageous to diagnose a disease as early as possible so that therapeutic drugs may be started and generally lead to reduced morbidity and mortality of that disease entity for that individual. A possible scenario would be if a patient could take a blood test to see if they have a biological marker for Rheumatoid Arthritis. If the marker were present, they could then seek treatment during the "subclinical" stage where they may only have a sensation of warmth in their joints instead of waiting until they have joint pain, swelling and deformity. That individual would likely have a much better long-term outcome for Rheumatoid Arthritis in comparison to someone who waits until they have a much later stage of the disease before seeking treatment.

The present invention is directed to the phenotyping of an organism or class or subclass of organisms. The phenotype is made up of data from a large number of data categories. The principle categories of data included within the scope of this invention are i) levels of cell populations including their cell associated molecules in biological fluid, ii) levels of soluble factors in the biological fluid, iii) drug dosing and pharmacokinetics (measurement of a drug and its metabolites in a body) and iv) clinical parameters. Additional categories of data may include, but are not limited to, i) levels of small molecule compounds in biological fluid, ii) genotype information regarding the individual, including the individual's genetic makeup and gene expression (mRNA or transcripts) levels, and iii) data obtained from assays of urine components. In certain embodiments data categories may include images such as x-ray, CAT scans of the brain or body, or MRIs, or information obtained from biopsies, EKGs, stress tests, endoscopies, ultrasound exams, laparascopic procedure, orthroscopic surgeries, PET scans, or any other measurement of an individual's condition.

In the preferred embodiment, the clinical parameters included in the database of the present invention would include, but not be limited to, the individual's age, gender, weight, height, body type, medical history (including comorbidities, medication, etc.), manifestations and categorization of disease or medical condition (if any) and other standard clinical observations made by a physician. Also included among the clinical parameters would be environmental and family history factors.

Clinical parameters could be further characterized by the source from which the information which is obtained. Patient obtained clinical parameters may include information that the patient provides via a questionnaire such as the WOMAC for

osteoarthritis, and the Health Assessment Questionnaire for Rheumatoid Arthritis which may be filled out in a doctor's office. Similarly, electronic or web-based questionnaires addressing all of a patient's current clinical symptoms could be completed by the patient prior to a clinic visit. Information obtained by a nurse would include vital signs, information from a variety of tests including allergy testing, pulmonary function testing, stress-thallium testing, or ECG tests. Clinical parameters collected from a physician includes a detailed history of prior illnesses, surgeries, hospitalizations, medications, reactions to medications, family history, social history, alcohol/drug/smoking history, as well as other behavior which would put a patient at high risk for HIV or Hepatitis. A thorough physical exam is also performed by a clinician and is a crucial component of a patient's clinical parameters.

In the preferred embodiment, the levels of cell populations and their associated molecules are identified by microvolume laser scanning cytometry. Such data can also be obtained by flow cytometry, but the volume of blood necessary to perform the flow cytometry assays places a serious limit on the number of assays that can be performed on blood taken from a given individual at one time. In addition, the sample preparation required for performing flow cytometry assays is time consuming, expensive, and may interfere with the measurement result.

The levels of soluble factors can be measured by any suitable technique. In the preferred embodiment, the levels of soluble factors is measured by standard immunoassay techniques, such as ELISA techniques. In an alternative embodiment, microvolume laser scanning cytometry is used to obtain levels of soluble factors. Soluble factors can be detected by immunoassays such as MLSC, ELISA, etc., mass spectrometry, 2D gel electrophoresis, combinations of mass spectrometry and immunosorption, and chemical assays. In the preferred embodiment, cell populations are detected by MLSC assays and soluble factors are detected by immunoassays or mass spectrometry.

The invention includes improved instrumentation for the rapid, reproducible and quantitative evaluation of biological parameters from a small quantity of blood; miniaturized, high sensitivity assays compatible with improved instrumentation for the detection of hundreds to thousands of biological parameters in blood; a broad clinical strategy to collect extensive medical information content from patients who are followed over time; software, databases and data mining tools to correlate patterns of parameters with normal biological functions, specific diseases, disease progression and response to

therapy; databases of clinical data and biological markers in collaboration with academic centers and clinical research institutes for use in drug development; development of diagnostic tests using proprietary patterns of markers and the ability to improve the efficiency of drug development by enabling more informed decisions in choosing lead compounds and identifying patients more likely to benefit from a given therapy.

The unique ability to phenotype an organism and to conduct reproducible and rapid measurements of large numbers of biological parameters is essential for the present invention to identify novel patterns of biological markers from small samples of blood. Statistical analyses to date have shown that the assays for the numbers of different cell subsets or cell populations, are quantitative and highly reproducible. The present technology, which uses small volumes of blood and requires limited handling of patient samples, has distinct advantages over other commercially available measurement technologies.

The invention further includes studies of patient populations related to particular diseases. These studies are based upon statistical analyses of disease patterns and require the collection of large numbers of blood samples from affected individuals. In addition, the present invention has utility for phenotyping and identifying biological markers in plants and animals and for assisting in preclinical studies.

Definitions

As used herein the term "phenotype" or "phenotyping" refers to a compilation comprising a substantial subset of all measurable characteristics of an organism. Such characteristics or parameters include, but are not limited to, levels of cell populations and their associated molecules, levels of soluble factors, levels of other molecules, genotype information, gene expression levels, genetic mutations, and clinical parameters. Such characteristics or parameters include all historical data and present data. For example, an organism's complete phenotype includes all measurable characteristics at the present time, as well as all such characteristics at all past points of time. In addition to technically measurable characteristics, the phenotype can include an organism's feelings or emotions (in the case where the organism is a human, the phenotype includes the individual's mental state, e.g., depression, pain, agitation, mental illnesses, chemical dependencies); diet and changes in diet, injuries, relational history, sexual practices, socio-economic status.

An used herein the term "organism" refers to all plants, animals, viruses and exoterrestial materials. Included within this definition, but not limited in any way, are humans, mice, rats, rabbits, companion animals, natural and genetically engineered plants, and natural and genetically engineered animals.

A given phenotype might include a compilation of characteristics of a single organism or a class or subclass of organisms. For example, the phenotypic data may be obtained from a single male individual who has been diagnosed with cancer before and after therapeutic intervention, a group of males between the age of 15 and 55, or a group of males between the ages of 15 and 55 diagnosed with cancer. In this manner, the phenotype may be specific to a given individual, or may represent the average or typical condition of a combined group of individuals.

The phenotype of an individual organism or group of organisms may be used for a variety of purposes. In the broadest scheme of the invention, the phenotype is looked at longitudinally and evaluated after some perturbation to the organism. For example, the comparison of the phenotype of an individual before and after exhibiting symptoms of asthma could be used to identify biological markers associated with asthma. In another example, the phenotype of an individual who has asthma can be compared with the phenotype of a population of normal adults. In another example, the phenotype of a naturally occurring plant can be compared with the phenotype of a genetically altered plant to determine what measurable characteristics are altered by the introduction of the genetic alteration. A further example of the use of phenotyipng information would be to periodically monitor well-patient status of an individual and to track measures of biological aging processes. The potential uses for comprehensive phenotypic data for an organism are almost infinite.

The present invention includes phenotypes for an organism or class or subclass of organisms, methods for obtaining such phenotypes and methods for utilizing such phenotypes, including for the identification of biological markers.

As used herein the term "biological marker" or "marker" or "biomarker" means a characteristic or parameter that is measured and evaluated as an indicator of normal and abnormal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention. Pharmacologic responses to therapeutic intervention include, but are not limited to, response to the intervention generally (e.g., efficacy), dose response to the intervention, side effect profiles of the intervention, and pharmacokinetic properties.

Response may be correlated with either efficacious or adverse (e.g., toxic) changes. Biological markers include patterns or ensembles of cells or molecules that change in association with a pathological process and have diagnostic and/or prognostic value.

Biological markers include, but are not limited to, levels of cell populations and their associated molecules, levels of soluble factors, levels of other molecules, gene expression levels (mRNA or transcripts), genetic mutations, and clinical parameters that can be correlated with the presence and progression of disease, normal biologic processes and response to therapy. Single biological markers currently used in both clinical practice and drug development include cholesterol, PSA, CD4 T cells, and viral RNA. Unlike the well known correlations between high cholesterol and heart disease, PSA and prostate cancer, and CD4 positive T cells and viral RNA and AIDS, the biological markers correlated with most other diseases have yet to be identified. As a result, although both government agencies and pharmaceutical companies are increasingly seeking development of biological markers for use in clinical trials, the use of biological markers in drug development has been limited to date.

As a non-limiting example, biological markers are often thought of as having discrete relationships with normal biological status, a disease or medical condition, e.g., high cholesterol correlates with an increased risk of heart disease, elevated PSA levels correlate with increased risk of prostate cancer, reduced CD4 T cells and increased viral RNA correlate with the presence/progression of AIDS. However, it is quite likely that useful markers for a variety of diseases or medical conditions may consist of significantly more complex patterns. For example, it could be discovered that lowered levels of one or more specific cell surface antigens on specific cell type(s) when found in conjunction with elevated levels of one or more soluble factors - - cytokines, perhaps - - is indicative of a particular auto-immune disease. Therefore, for the purposes of this invention, a biological marker may refer to a pattern of a number of indicators.

As used herein the term "biological marker identification system" means a system for obtaining information from a patient population and assimilating the information in a manner that enables the correlation of the data and the identification of biological markers. A biological marker identification system comprises an integrated database comprising a plurality of data categories, data from a plurality of individuals corresponding to each of said data categories, and processing means for correlating data within the data categories, wherein correlation analysis of data categories can be made to identify the data category or

categories where individuals having said disease or medical condition may be differentiated from those individuals not having said disease or medical condition, wherein said identified category or categories are markers for said disease or medical condition. Additionally, markers may be identified by comparing data in various data categories for a single individual at different points of time, e.g., before and after the administration of a drug.

As used herein the term "data category" means any type of measurement that can be discerned about an organism. Examples of data categories useful in the present invention include, but are not limited to, numbers and types of cell populations and their associated molecules in the biological fluid of an organism, numbers and types of soluble factors in the biological fluid of an organism, information associated with a clinical parameter of an organism, cell volumetric counts per ml of biological fluid of an organism, numbers and types of small molecules in the biological fluid of an individual, genomic information associated with the DNA of an organism and gene expression levels. For example, a single data category would represent the concentration of IL-1 in the blood of an organism. Additionally, a data category could be the level of a drug or its metabolites in blood or urine. An additional example of a data category would be absolute CD4 T cell count. The number or information assigned to an organism or class or subclass of organisms at any given point in time in part comprises the phenotype of that organism.

As used herein the term "biological fluid" means any biological substance, including but not limited to, blood (including whole blood, leukocytes prepared by lysis of red blood cells, peripheral blood mononuclear cells, plasma, and serum), sputum, urine, semen, cerebrospinal fluid, bronchial aspirate, sweat, feces, synovial fluid and whole or manipulated tissue. Biological fluid typically contains cells and their associated molecules, soluble factors, small molecules and other substances. Blood is the preferred biological fluid in this invention for a number of reasons. First, it is readily available and can be drawn at multiple times. Blood replenishes, in part, from progenitors in the marrow over time. Blood is responsive to antigenic challenges and has a memory of antigenic challenges. Blood is centrally located, recirculates and potentially reports on changes throughout the body. Blood contains numerous cell populations, including surface molecules, internal molecules, and secreted molecules associated with individual cells. Blood also contains soluble factors that are both self, such as cytokines, antibodies, acute phase proteins, etc., and foreign, such as chemicals and products of infectious diseases.

As used herein the term "cell population" means a set of cells with common characteristics. The characteristics may include the presence and level of one, two, three or more cell associated molecules, size, etc. One, two or more cell associated molecules can define a cell population. In general some additional cell associated molecules can be used to further subset a cell population. A cell population is identified at the population level and not at the protein level. A cell population can be defined by one, two or more molecules. Any cell population is a potential marker.

As used herein the term "cell associated molecule" means any molecule associated with a cell. This includes, but is not limited to: 1) intrinsic cell surface molecules such as proteins, glycoproteins, lipids, and glycolipids; 2) extrinsic cell surface molecules such as cytokines bound to their receptors, immunoglobulin bound to Fc receptors, foreign antigen bound to B cell or T cell receptors and auto-antibodies bound to self antigens; 3) intrinsic internal molecules such as cytoplasmic proteins, carbohydrates, lipids and mRNA, and nuclear protein and DNA (including genomic and somatic nucleic acids); and 4) extrinsic internal molecules such as viral proteins and nucleic acid. The preferred cell associated molecule is typically a cell surface protein. As an example, there are hundreds of leukocyte cell surface proteins or antigens, including leukocyte differentiation antigens (including CD antigens, currently through CD166, see, Leucocyte Typing VI, Kishimoto, T. et al.. ED, 1997), antigen receptors (such as the B cell receptor and the T cell receptor), and major histocompatibility complex. Each of these classes encompass a vast number of proteins. A list of exemplary cell surface proteins is provided in Table 1, which is merely an illustration of the vast number of cell surface proteins and is in no way intended to be a comprehensive list.

As used herein the term "soluble factor" means any measurable component of a biological fluid or tissue that is not a cell population or cell associated molecule. Soluble factor includes, but is not limited to, soluble proteins, carbohydrates, lipids, lipoproteins, steroids, other small molecules, including metallic, inorganic, ionic and metallorganic species and complexes of any of the preceding components, e.g., cytokines and soluble receptor; antibodies and antigens; and a drug complexed to anything. Soluble factors can be both self, such as cytokines, antibodies, acute phase proteins, etc., and foreign, such as chemicals, products of infectious diseases and intestinal flora and fauna. Soluble factors may be intrinsic, i.e., produced by the organism, or extrinsic such as a virus, drug or environmental toxin. Soluble factors can be small molecule compounds such as

prostaglandins, vitamins, metabolites (such as iron, sugars, amino acids, etc.), drugs and drug metabolites. A list of exemplary soluble proteins is provided in Table 6, which is merely an illustration of the vast number of soluble proteins and is in no way intended to be a comprehensive list.

For the purposes of this invention, soluble factors may be either known or unknown entities. A variety of techniques are available where a given species may be identifiable, but the chemical identity of the species is unknown. In the present invention, the chemical identity of the soluble factor need not be currently known or known at the time the assay is performed to determine its presence or absence.

As used herein the term "small molecule" or "organic molecule" or "small organic molecule" means a soluble factor or cell associated factor having a molecular weight in the range of 18 to 10,000. Small molecules can include, but are not limited to, prostaglandins, vitamins, metabolites (such as iron, sugars, amino acids, etc.), drugs and drug metabolites.

As used herein the term “disease or medical condition” means an interruption, cessation, disorder or change of body functions, systems or organs. Examples of disease or medical conditions include, but are not limited to, immune and inflammatory conditions, cancer, cardiovascular disease, infectious diseases, psychiatric conditions, obesity, and other such diseases. By way of illustration, immune and inflammatory conditions include autoimmune diseases, which further include rheumatoid arthritis (RA), multiple sclerosis (MS), diabetes, etc.

As used herein the term "perturbation" means an exterior or interior measurable event that can occur to an organism. A simple example would be the administration of a therapeutic agent to an individual, or an individual that was healthy and then developed asthma. In this application a perturbation may also include differences between an individual or groups of organisms that are being compared. For example, a population of animals may be considered to be normal, and their phenotype is being compared to the phenotype of a similar but genetically altered animal. The individual genetically altered animal was perturbed in the sense that its genetic alteration was perturbed from normal. In many cases the perturbation is not a single event that occurs at a discrete point in time. The perturbation may occur over an extended period of time, and/or may be cyclical or intermittent.

As used herein the term “clinical parameter” means information that is obtained that may be relevant to a disease or medical condition. Such information may be supplied by

the patient or by a medical or scientific observer. Examples of clinical parameters for humans include, but are not limited to, age, gender, weight, height, body type, medical history, ethnicity, family history, genetic factors, environmental factors, manifestation and categorization of disease or medical condition, and any result of a clinical lab test, such as blood pressure, MRI, x-ray, etc.

Clinical parameters could be further characterized by the source of information which is obtained. Patient obtained clinical parameters may include information that the patient provides via a questionnaire such as the WOMAC for osteoarthritis, and the Health Assessment Questionnaire for Rheumatoid Arthritis which may be filled out on paper in a doctor's office. Similarly, an electronic or web-based questionnaires addressing all of a patient's current clinical symptoms could be completed by the patient prior to a clinic visit. Information obtained by a nurse would include vital signs, information from a variety of tests including allergy testing, pulmonary function testing, stress-thallium testing, or ECG tests. Clinical parameters collected from a physician includes a detailed history of prior illnesses, surgeries, hospitalizations, medications, reactions to medications, family history, social history, alcohol/drug/smoking history, as well as other behavior which would put a patient at high risk for HIV or Hepatitis. A thorough physical exam is also performed by a clinician and is a crucial component of a patient's clinical parameters.

As used herein the term "genotype information" means any data relating to the organisms genetic makeup, gene mutations, gene expression, e.g., mRNA or transcription levels, and any other measure or parameter associated with the genetic material of the organism.

As used herein the term "clinical endpoint" means a characteristic or variable that measures how a patient feels, functions, or survives. There are several mechanism which are commonly used to measure how a patient feels or functions with a specific disease and they often include validated clinical questionnaires. These may be self administered such as the Beck's depression questionnaire or the International Prostate Questionnaire to determine if changes in urination are due to prostatic hypertrophy v. bladder outlet obstruction. These tools may be given by a health care provider who is judging features such as facial expression, inability of patient to sit down for more than 10 minutes, level of agitation etc., while completing the Carrol Questionnaire to determine if a patient is manic. And finally, in the case of psychiatric illness, typically patient's who are admitted for a hospitalization for an acute exacerbation of their illness will be observed without realizing

it by a clinician to note their ability to function in a variety of settings including group interactions or making lunch. These "clinical endpoints" are highly variable per disease entity and subsequently the tools which are used to characterize these endpoints are also quite broad.

As used herein the term "Microvolume Laser Scanning Cytometry" or "MLSC" means a method for detecting the presence of a component in a small volume of a sample using a fluorescently labeled detection molecule and subjecting the sample to optical scanning where the fluorescence emission is recorded. The MLSC system has several key features that distinguish it from other technologies: 1) only small amounts of blood (5-50 μl) are required for many assays; 2) absolute cell counts (cells/ μl) are obtained; and, 3) the assay can be executed either directly on whole blood or on purified white blood cells. Implementation of this technology will facilitate measurement of several hundred different cell populations from a single harvesting of blood. The MLSC technology is described in United States Patent Numbers 5,547,849 and 5,556,764 and in Dietz et al. (Cytometry 23:177-186 (1996)), and provisional patent application entitled "Laser-Scanner Confocal Time-Resolved Fluorescence Spectroscopy System" (United States Provisional Application Number 60/144,798, filed July 21, 1999), and the commonly-owned utility application filed concurrently with the present application, entitled "System for Microvolume Laser Scanning Cytometry", each of which is incorporated herein in its entirety. Laser scanning cytometry with microvolume capillaries provides a powerful method for monitoring fluorescently labeled cells in whole blood, processed blood, and other fluids. The present invention further improves MLSC technology by improving the capacity of the MLSC instrument to do simultaneous measurement of multiple biological markers from a small quantity of blood. A schematic of the improved SurroScan optical system is shown in Figure 2.

As used herein the term "tag" means any entity or species, including but not limited to an atom, a molecule, a fragment of molecule or a functional group; a particle or combination of particles; a single or sequence of electromagnetic pulses; or any other form of matter associated with, attached to (either covalently or non-covalently), or otherwise connected to a component of a biological system (a molecule or collection of molecules such as cell, a cation, an anion, an atom, or any supramolecular assembly, including but not limited to non-covalent complexes between biological molecules) that is used to, identify,

quantify, associate, recognize, follow, spot, make out, see, name, track, or otherwise distinguish (henceforth I/Q) said component.

Tags are often extrinsic, i.e. not part of the component under investigation. For example, a fluorescent dye molecule is often used as a tag, either for tracking, quantitation, or both. Likewise, the use of biotin or streptavidin as tag, linked to a secondary species such as an enzyme for ELISA, is widespread. Other forms of tags include, but are not limited to, isotopic mass tags for protein I/Q by mass spectrometry, Raman-active tags for I/Q by Raman scattering, particulate tags for I/Q by light scattering, fluorescence, agglutination, energy transfer, and a variety of other detection mechanisms, including surface plasmon resonance.

In this regard, there are almost an infinite number of particulate tags, only a small number of which have been previously used. As nanoparticle science is in its infancy (as was organic chemistry two centuries ago), one can anticipate that the complexity of particulate tags will approach molecular complexity. In other words, we expect that particulate tags might rival the organic molecules currently used as bead tags in combinatorial chemistry, in other words thousands to hundreds of thousands or even millions of uniquely identifiable tags. We further anticipate that such tags will become small enough to allow all intracellular measurements. For example, there are now roughly one-half dozen different luminescent semiconducting quantum dot nanoparticles, each fluorescing at a different wavelength. In theory, one could anticipate production of thousands or millions of such orthogonal nanoparticulate optical tags, although the detection mechanism may or may not involve fluorescence (or even other optical methods).

The same could be said of supramolecular science, and supramolecular tags. We anticipate that molecular assemblies held together by non-covalent forces could ultimately find use as tags. Furthermore, tags could comprise individual molecules either covalently or non-covalently associated with biological components. For example, one could imagine using electrochemically-active redox tags to uniquely identify components. If one had 10 different molecules, each with a different redox potential, and each pre-functionalized to react with a particular biological component, then one could carry out multiplexed tag I/Q, using the detection of the redox potential as the identifying characteristic. This is identical to the strategy currently used with fluorescence, with redox "space" used in lieu of "wavelength" space.

Note also that a tag can be a functional group, as in a carboxylate, an amine, a sugar, etc., or even a spin associated with a molecule. For example, we anticipate the possibility that two samples could be mixed together, with each sample having one or more nuclei imparted with a particular sequence of electromagnetic pulses (of the sort typically used in high-field NMR). We further envision that the pulses for two samples would be long-lived enough to compare them using a method of detection. In particular, we envision that possibility that the signatures for the two samples would cancel for all species where the concentrations are identical, leaving behind a signal only for those species where concentrations in the two samples are non-identical.

It should be clear to a person skilled in the art that there is no functional difference between tags, as defined above, and "reporters" or "reporter molecules", as typically used in the chemical and biological literature. Likewise, a "detection molecule" as defined below, can itself be a tag (for example when I/Q is based on mass, as in quartz crystal microbalances or piezo inertial biosensors).

As used herein the term "detection molecule" means any molecule or molecular assembly capable of binding to a molecule or other species of interest, including but not limited to a cell-associated molecule, a soluble factor, or a small molecule or organic molecule. Preferred detection molecules are antibodies. The antibodies can be monoclonal or polyclonal. Note, however, that as new types of detection molecules are discovered and popularized, they certainly can be used. For example, aptamers are increasingly being used for molecular recognition, and organic chemists have now synthesized a large number of molecular receptors. Ultimately, these could be used as detection molecules, either by themselves or in association with a tag.

As used herein the terms "dye", "fluorophore", "fluorescent dye" are used interchangeably to mean a molecule capable of fluorescing under excitation by a laser. The dye is typically directly linked to a detection molecule in the present invention, although indirect linkage is also encompassed herein. Many dyes are well known in the art and include, but are not limited to those shown in Table 2. In certain preferred embodiments, fluorophores are used which can be excited in the red region (> 600 nm) of the spectrum. Two red dyes, Cy5 and Cy5.5, are typically used. They have emission peaks of 665 and 695 nanometers, respectively, and can be readily coupled to antibodies. Both can be excited at 633 nm with a helium-neon laser. Sets of 3 red dyes that may be used include, Cy5, Cy5.5 and Cy 7 or Cy5, Cy5.5 and Cy 7 APC. See, Mujumdar <u>et al.</u>, Bioconjugate

Chemistry, 7:356 (1996); United States Patent No. 5,268,486; Beavis et al., Cytometry, 24:390 (1996); Roderser et al., Cytometry, 24:191 (1996); and United States Patent No. 5,714,386. Additional novel dyes useful for tagging or detection purposes within the present invention are described in commonly-owned United States Provisional Application Number 60/142,477, filed July 6, 1999, entitled "Bridged Fluorescent Dyes, Their Preparation and Their Use in Assays," incorporated herein in its entirety by this reference.

As used herein the term "animal model" refers to any experimental animal system in which diseases or conditions with similar pathology and progression to human diseases or medical conditions can be developed. Suitable animal systems include, but are not limited to, rats, mice, rabbits, and primates. In some cases, the disease arises spontaneously in the animal model. In other cases, the induction of disease in the animal model can result from exposure to the same conditions--for example, infection with a pathogen, exposure to a toxin, or a particular diet--that causes the disease in humans. Alternatively, the disease or condition can be induced in the animal model with agents that mimic the human disease or medical condition even if the actual initiator(s) of the human disease or medical condition is unknown. The disease or medical condition might also be induced through the use of surgical techniques. Genetic manipulation of experimental animal model systems provides a further tool for the development of the animal models, either standing alone or in combination with the other methods of disease induction.

Preclinical Applications of Phenotyping

Currently much effort is being directed towards the identification and analysis of biological markers in humans. However, it would be desirable to have a method for identifying and analyzing biological markers in experimental animal systems. For example, biological markers of the progression of a particular human disease could be identified in an experimentally-induced animal model of that disease, *e.g.*, the rat adjuvant model of arthritis (reviewed in Philippe, *et al*, American Journal of Physiology 273:R1550-56 (1997)). Using the identified markers, the efficacy of experimental therapeutics could be determined in the animal model. Therapeutics that have a highly specific effect on the expression of biological markers in animals, which markers are prognostic or diagnostic of the same disease in humans, can therefore be identified without conducting early--and hence risky--human clinical trials. Alternatively, novel biological markers can be identified in experimental animal models of human disease, and then experiments can be

performed to determine whether the same markers, or their human homologues, are prognostic or diagnostic of the same disease or medical condition in humans. In some cases, biological markers identified in humans can be used to facilitate preclinical trials where animal models can be evaluated by the corresponding biological markers. The present invention provides methods and instrumentation for performing such analyses.

In one series of embodiments of the invention, the expression of biological markers is studied in an animal model of a human disease. There are currently many such models and many more being developed, using a variety of different techniques to induce the specific disease. In each case, the biological markers of interest can be initially identified in preferred embodiments using MLSC. The identified markers can then be studied using MLSC to determine the response of the animal to a candidate therapeutic. Because MLSC-based assays typically only require small volumes of biological fluid, MLSC is uniquely suited for use in animal model systems (especially in rat and mouse) where only limited amounts of fluid can be obtained from an animal without sacrificing it. In particular, the use of MLSC will permit multiple time point analysis of an experimental animal to determine the pharmacokinetics of a candidate therapeutic.

In some embodiments of the invention, the animal homologues of known or newly identified human biological markers of a particular disease are studied in an experimentally-induced animal model of that disease. In many cases, the animal homologues of human molecules will already be known and characterized. For example, through extensive study, a great deal is known about proteins that behave similarly in mouse and in humans. The identification of previously unknown animal homologues of human biological markers, and the preparation of reagents that can bind to them, can be accomplished through the use of standard molecular biology techniques well known in the art.

In other embodiments of the invention, novel biological markers--for example, a previously unknown pattern of expression of known blood cell-associated proteins--may be initially observed in an animal model of a human disease. In this embodiment, the relevancy of the identified markers to the progression or development of the human disease can be determined by identifying human homologues of the biological markers, and then studying their expression in humans suffering from the disease of interest. If the identified animal biological markers appear to be relevant to the human disease, then they can serve: 1) as the basis of new diagnostic and prognostic assays for the disease in humans; and 2)

as a means for evaluating the specificity and efficacy of candidate therapeutics in the animal model of the disease.

In one embodiment of the present invention new and improved animal models may be developed based on biological markers identified in humans. For example, utilizing the biological marker identification system of the present invention it can be found that for a given disease or medical condition that the level of given soluble factor in serum is greatly increased, while the level of certain cell population is decreased. Based on this information, animal models can be tailored -- for example by the use of genetic knockouts of homologous factors -- to better simulate the disease in the animal serum.

The phenotyping system of the present invention may also be useful in the identification of new or improved animal models. For example, by phenotyping a number of genetically altered animals, a fuller picture of the manifestations of the genetic alternations can be recognized. Utilization of this knowledge can be useful in identifying new or improved animal models. For example, it may be possible to create a number of genetic knock-out mice that all appear to simulate a chosen human disease state. However, by phenotyping each of the various knock-outs, as well as humans that suffer from the disease, it will be possible to identify the animal model that most closely mimics the human disease.

The present invention can be used in any animal model of a human disease. By way of illustration only, the present invention can be used to identify and analyze biological markers in animal models of many aspects of cardiovascular disease, including hypertension, artherosclerosis, cardiac hypertrophy, atherogenesis, and thrombosis. Many animal models of congestive heart failure and hypertrophy are currently being developed, and a number are reviewed in: Carmeliet, Artherosclerosis, 144:163-93 (1999); Young et al., Molecular Basis of Cardiovascular Disease, 37-85 (K.R. Chien, Editor) (1999); Hasenfuss, Cardiovasc. Res. 39:60-76 (1998); Krege et al, Fundam. Clin. Cardiol. 26:271-92 (1996); Liao et al., Am. J. Therap. 4:149-58 (1997); and Becker et al., Hypertension 27:495-501 (1996) The following is a partial list of some animal models of cardiovascular disease:

- The JCR:LA-cp rat model of human vascular disease can be used to identify and study biomarkers that correlate with insulin resistance, vasculopathy, and cardiovascular disease. O'Brien et al. Can. J. Physiol. Pharmacol. 76: 72-76 (1998).

- Animal models of insulin-dependent diabetes have been used to study the development of ischemic heart disease in the diabetic population. Reviewed in: Pierce et al., Can. J. Physiol. 75:343-50 (1997).
- Infection of mouse, rabbits and monkeys with Chlamydia pneumonia has been used to investigate that pathogens role in the development of asthma and cardiovascular disease in humans, as reviewed in: Saikku et al, Artherosclerosis, 140 (Suppl. 1), S17-S19 (1998).
- Spontaneously hypertensive rat (SHR) strains, and SHR strains carrying a portion of chromosome 13 (including the renin gene) from normotensive rats (SHR.BN-Ren) can be used to investigate the interaction between high blood pressure and dyslipidemia in cardiovascular disease. St. Lezin et al, Hypertension, 31:373-377 (1998).
- Spontaneously occurring hypertrophic cardiomyopathy in Landrace pigs may be a useful model of cardiovascular disease in humans. Chiu et al., Cardiovasc. Pathol. 8:169-75 (1999).
- Cardiomyopathic hamster strains can be used to investigate the role of brain and atrial natriuretic peptides (BNP and ANP) in human cardiovascular disease. Tamura et al., J. Clin. Invest. 94:1059-68 (1994).
- Hypertensive and atherogenic rat strains have been used as models for the study of the effect of dietary salt, protein and lipids on the pathogenesis of human cardiovascular disease. Reviewed in: Yamori et al., Nutritional Prevention of Cardiovascular Disease (Symposium Proceedings) (1984), Published by: Academic Press, Orlando, Fla.
- Rat, guinea pig, rabbit, dog, sheep, and baboon models of preeclampsia have been used to study the pathophysiology this hypertensive disorder of human pregnancy. Reviewed in: Hypertension in Pregnancy 12:413-37 (1993).

In other embodiments, the present invention is used to identify and analyze biological markers in animal models of inflammatory diseases such as arthritis and multiple sclerosis.

When used to screen candidate therapeutics, the present invention has a number of significant advantages over more traditional screening methodologies. Firstly, clinical testing comes at a relatively late stage in the development of the therapeutic, at which point the therapeutic is known to have a highly specific effect on the expression of analogous animal biological markers; this minimizes the risks to the clinical participants. Secondly, using experimental animal models to analyze patterns of biological marker expression

means that only relatively small quantities of the potential therapeutic need be synthesized initially, thus reducing the cost of therapeutic development.

In other embodiments, the methods and systems of the present invention are used to identify markers of disease or medical conditions in animals for veterinary purposes. The identified markers can then be used to screen for candidate therapeutics directed against that disease or condition. This embodiment can be applied to domesticated animals, livestock and plants.

Instrumentation

Any suitable means for obtaining data that meet the requirements of the data categories is within the scope of the invention. In the preferred embodiments, Microvolume Laser Scanning Cytometry ("MLSC") is used to obtain the data for cell associated molecules and cell type count. In some embodiments, the MLSC technology is used with a bead based capture system or with various types of enzyme linked immunosorbent assays (such as ELISA) to obtain data for soluble proteins. Another preferred means for obtaining data for compounds, particularly small molecules, includes the use of mass spectrometry. The MLSC technology used in this invention, is a powerful method for monitoring fluorescently labeled cells and soluble proteins in blood. This technology is currently used in clinical laboratories for the identification of one or two cellular markers for diagnostic applications. The present invention uses MLSC to facilitate the identification of biological parameters. In one embodiment, the present invention improves MLSC technology by improving the capacity of the MLSC instrument to do simultaneous measurement of multiple biological characteristics or parameters from a small quantity of blood.

Specific enhancements achieved with the instrument of the invention (termed "SurroScan instrument") include the following: 1) two additional fluorescence color channels allow simultaneous detection and measurement of up to four fluorescent colors; 2) higher laser excitation power improves sensitivity and throughput; 3) disposable capillary arrays allow more assays per patient sample using less blood per assay; 4) improved software and system integration automates sample measurements and data analysis; 5) the capacity of SurroScan instruments is expanded to handle higher volumes of patient samples for database creation and biological marker discovery.

Microvolume Laser Scanning Cytometry (MLSC) System Design

The MLSC technology is described in United States Patent Numbers 5,547,849 and 5,556,764 and in Dietz et al. (Cytometry 23:177-186 (1996)), each of which is incorporated herein in its entirety. The Imagn 2000 system, commercially available from Biometric Imaging Inc., is an example of a MLSC system. Laser scanning cytometry with microvolume capillaries provides a powerful method for monitoring fluorescently labeled cells in whole blood, processed blood, and other fluids. The present invention further improves MLSC technology by improving the capacity of the MLSC instrument to do simultaneous measurement of multiple biological markers from a small quantity of blood. A schematic of the improved SurroScan optical system is shown in Figure 2. The preferred MLSC instrument for use in the present invention is described in commonly owned United States Provisional Application No. 60/144,798, filed July 21, 1999, entitled "System for Microvolume Laser Scanning Cytometry" and in the commonly-owned utility application filed concurrently with the present invention entitled "System for Microvolume Laser Scanning Cytometry". Both of these applications are incorporated herein by reference in their entirety.

One embodiment of the improved optical configuration is shown in Figure 2. A capillary array 10 contains samples for analysis. In the preferred embodiment, collimated excitation light is provided by one or more lasers. In particularly preferred embodiments, excitation light of 633nm is provided by a He-Ne laser 11. This wavelength avoids problems associated with the autofluorescence of biological materials. The power of the laser is increased from 3 to 17 mW. Higher laser power has two potential advantages, increased sensitivity and increased scanning speed. The collimated laser light is deflected by an excitation dichroic filter 12. Upon reflection, the light is incident on a galvanometer-driven scan mirror 13. The scan mirror can be rapidly oscillated over a fixed range of angles by the galvanometer e.g. +/- 2.5 degrees. The scanning mirror reflects the incident light into two relay lenses 14 and 15 that image the scan mirror onto the entrance pupil of the microscope objective 16. This optical configuration converts a specific scanned angle at the mirror to a specific field position at the focus of the microscope objective. The +/- degree angular sweep results in a 1 mm scan width at the objective's focus. The relationship between the scan angle and the field position is essentially linear in this configuration and over this range of angles. Furthermore the microscope objective focuses

the incoming collimated beam to a spot at the objective's focus plane. The spot diameter, which sets the optical resolution, is determined by the diameter of the collimated beam and the focal length of the objective.

Fluorescence samples placed in the path of the swept excitation beam emit stokes-shifted light. This light is collected by the objective and collimated. This collimated light emerges from the two relay lenses 14 and 15 still collimated and impinges upon the scan mirror which reflects and descans it. The stokes-shifted light then passes through a dichroic excitation filter (which reflects shorter wavelength light and allows longer wavelength light to pass through) and then through first long pass filter 17 that further serves to filter out any reflected excitation light.

The improved instrument of the instant invention then uses a series of further dichroic filters to separate the stokes-shifted light into four different emission bands. A first fluorescence dichroic 18 divides the two bluest fluorescence colors from the two reddest. The two bluest colors are then focussed onto first aperture 19 via a first focusing lens 20 in order to significantly reduce any out-of-focus fluorescence signal. After passing though the aperture, a second fluorescence dichroic 21 further separates the individual blue colors from one another. The individual blue colors are then parsed to two separate photomultipliers 22 and 23. The two reddest colors are focused onto a second aperture 24 via a second long pass filter 25, a mirror 26, and a second focusing lens 27 after being divided from the two bluest colors by first fluorescence dichroic 28. After passing through aperture 24, the reddest colors are separated from one another by third fluorescence dichroic 28. The individual red colors are then parsed to photomultipliers 29 and 30. In this way, four separate fluorescence signals can be simultaneously transmitted from the sample held in the capillary to individual photomultipliers. This improvement, for the first time, allows four separate analytes to be monitored simultaneously. Each photomultiplier generates an electronic current in response to the incoming fluorescence photon flux. These individual currents are converted to separate voltages by one or more preamplifiers in the detection electronics. The voltages are sampled at regular intervals by an analog to digital converter in order to determine pixel intensity values for the scanned image. The four channels of the instant invention are named channel 0, 1, 2, and 3.

The new optical layout has four detection channels to allow simultaneous measurement of up to 4 fluorescently labeled molecules. In a preferred embodiment, multiple-color assays are used. Typically 3 or more fluorescent colors are used in each

assay. Under circumstances where appropriate dye combinations are available, the instrument is capable of supporting 4-color assays.

An XY translational stage is used to move an array of capillaries relative to the optical system. The SurroScan system translation stage holds two arrays, each of which has the footprint of a 96-well plate. Capillary arrays have been designed which have 32 fixed capillaries each and spacing that is compatible with multi-channel pipettes. The operator is able to load two plates of 32 capillaries at a time. No operator intervention is needed while the plates are scanned and the images are processed. As an alternative, 16 individual capillaries designed for the Imagn 2000 (VC120) are loaded into alternative holders.

Image processing software accommodates images with either 2, 3, or 4 colors of fluorescent dyes. The software automatically identifies and parameterizes particles detected in any of the individual colors. The measured parameters describing each particle are saved in a list-mode format, which is made compatible with conventional cytometry analysis software, such as FlowJo.

A new disposable cartridge design containing arrays of capillaries has been developed and is described in Provisional United States Patent applications, (United States Provisional Application Number 60/130,876, entitled, "Disposable Optical Cuvette Cartridge", filed April 23, 1999; United States Provisional Application Number 60/130,918, entitled "Spectrophotometric Analysis System Employing a Disposable Optical Cuvette Cartridge", filed April 23, 1999; and United States Provisional Application Number 60/130,875, entitled "Vacuum Chuck for Thin Film Optical Cuvette Cartridge", filed April 23, 1999), and the commonly-owned utility application filed April 20, 2000, entitled "Disposal Optical Cuvette Cartridge", which are incorporated by reference herein in their entirety. This capillary cartridge is used in Examples 5, 7 and 8. The design currently in use, called Flex-32, contains 32 capillaries. Fill holes in the FLEX32-plates have the same 9 mm spacing as 96-well plates and multichannel pipetting devices. It is constructed from 2 layers of mylar sandwiched together with a double-sticky adhesive layer which is die-cut to define the capillary inner dimensions. The resulting cartridge can be manufactured at low cost in high volumes. The cartridge is flexible, which allows it to be held onto an optically flat baseplate by vacuum pressure, removing the requirements for flatness in the manufacturing process. The capillary spacing was designed to retain compatibility with multi-channel microplate pipetters and robotics.

Cellular Assays

The invention includes cellular assays, many of which are antibody based, that are compatible with instrumentation, preferably MLSC instrumentation and are capable of measuring hundreds to thousands of cell populations and their cell associated molecules from a single 10 mL tube of blood. In one preferred embodiment, any type of detection molecule and assay format compatible with MLSC is encompassed in this invention, including, but not limited to cell surface proteins including markers of activation and adhesion, intracellular molecules, assays to distinguish changes in activation states of cells, assays to concentrate and identify rare white cells, assays for use with whole blood, and assays for detection of soluble factors, such as proteins, in blood.

As with flow cytometry, fluorophore-labeled antibodies specific for cell surface antigens are used to identify, characterize and enumerate specific populations. The reaction can be done in whole blood. In general, there is no need to wash the reagent away; quantitative dilution of the blood-antibody mixture is usually sufficient sample preparation. The cell-antibody mixture is loaded into an optical-quality capillary of known volume and analyzed with a laser-based fluorescence imaging instrument. In order to operate with whole blood, fluorophores are used which can be excited in the red region (> 600 nm) of the spectrum. Purified white blood cells can also be analyzed with the instrument. In contrast to flow cytometry, the laser scans over stationary cells rather than cells flowing past the laser. A small cylindrical laser spot is scanned across the capillary in one direction while the capillary is translated relative to the optical system in a second direction. Photomultiplier tubes are used to detect the fluorescent signal. Image-processing software is used to analyze the image and identify and enumerate the cells of interest.

This MLSC approach allows one to obtain absolute cell counts on hundreds of different cell populations from a single tube of blood. For a set of antibodies to 100 different antigens, there are about 5000 possible 2-color combinations and about 162,000 possible 3-color combinations. (n combinations r at a time, ${}_nC_r = n!/r!(n-r)!$) so careful thought is needed to develop the most appropriate set of 100 or so assays. Multi-color capability allows more cell populations to be identified with a given amount of blood than the original 2-color system. As an example, by multiplexing reagents all populations identified in two 2-color assays can be identified in one 3-color assay. For example it is possible to assay CD3, CD4 and CD8 in one capillary instead of CD3 and CD4 in one

capillary and CD3 and CD8 in another. More importantly, unique cell populations can be defined by the simultaneous expression of three or more antigens. For example, CD8 T cells can be subsetted into 4 different populations based on the differential expression of CD45RA and CD62L.

Immunoassay Procedures

Immunoassays can be run in a variety of formats and any appropriate format is envisioned in the present invention. Two examples are given below. The MLSC system can be used with microsphere-based immunoassays. In this sandwich assay, the microsphere is used as a solid support for an analyte-specific capture antibody. Analyte from a biological fluid is bound to the antibody-coated microsphere and detected with a second antibody, which is directly labeled with a fluorescent molecule such as Cy5, and which binds to a distinct epitope on the analyte. A protocol using amino beads and a heterobifunctional crosslinker to covalently attach antibodies via their hinge region works well in multiple assays. It is possible to distinguish beads of different sizes (3 to 20 micron range) with the MLSC instrument and current software. By coupling different capture antibodies to microspheres of different sizes it is possible to multiplex immunoassays in a single capillary. Internal indicator dyes can also be used to distinguish microspheres and facilitate multiplexing.

The immunoassays for soluble factors discussed in Example 5 are all chemiluminescent – based sandwich ELISA. Microtiter plates are coated with capture antibodies specific for the analyte of interest and blocked. Biological fluid containing the analyte is added, incubated and then washed. Biotinylated antibody specific for a second epitope on the same analyte is added, incubated and washed followed by an avidin-alkaline phosphatase conjugate. The level of analyte is revealed with a chemiluminescent alkaline phosphatase substrate. Plates are read in a Wallac Victor2 luminometer or similar instrument.

Design and implementation for a robust panel of cellular assays to aid the discovery of biological markers for diseases or medical conditions

The MLSC system is designed to allow rapid staining of cells using minimal quantities of blood. Reagents directed against scores of different cell surface antigens are

developed, which when combined can identify hundreds of different cell populations. The strategy for reagent and combination development is discussed below.

A set of monoclonal antibody reagents are employed which are suitable for developing more than 100 cellular assays. To date, many (about 120) different monoclonal antibodies directed against numerous (about 80) different cell surface antigens have been successfully identified and tested with the 2-color MLSC instrument. The small organic dyes like Cy5 and Cy5.5 are readily coupled to the amino groups of antibodies using single-step NHS chemistry and well established procedures. Preferred dye-to-antibody ratios, have been determined for Cy5, Cy5.5, and Cy7 reagents, and are generally in the range of one to four. Protein fluorochromes, like APC, are linked to the sulfhydryl groups of moderately reduced antibody in a 3-step procedure using the heterobifuntional crosslinking reagent SMCC. Preparation of reagents containing other fluorophores is also possible. The preparation of Cy7-APC and (Cy7-APC)-antibody conjugates for flow cytometry applications has been previously described. The antibody-fluorophore coupling chemistry is the same as for APC. All protein-protein conjugates are purified by traditional means, such as, by gel filtration on an Akta FPLC. Fluorescent microspheres can also be investigated. Antibodies are coupled with 2-step carbodiamide chemistry to carboxylated microspheres.

New monoclonal antibodies reagents are titrated on both whole blood and lysed red blood cells. Reagent specificity, and lack of non-specific binding, is confirmed with appropriate counter stains. Analysis is done with any appropriate software program, including FlowJo cytometry software (Treestar, Inc available as an Internet download at http://www.treestar.com/ flowjo/). From the titration the optimal amount of each reagent per assay (typically 0.01 to 2 μg/ml) and preliminary analysis criteria is determined. In the preferred embodiment, all assays are conducted in homogenous (no wash) mode. This generally requires that each antibody reagent have a titer point of ≤1 μg/ml so that the fluorescence background is not too high. A potential difficulty may be that a particular reagent may not be amenable to conjugation or may have too high of a titer point. It is usually possible to substitute a second monoclonal antibody to the same antigen. There is also a risk that some individual antigens may not be measurable during the time course of the study. Multiple antibodies from each antigen category are typically evaluated.

Typically, a panel of about 50-100 (or greater) cellular assays is developed for monitoring a disease or medical condition. Such assays enable one to enumerate hundreds of different cell populations. As an example, it is possible to monitor the immune and inflammatory cellular parameters potentially significant for rheumatoid arthritis (RA). For RA populations, the cell surface antigens being evaluated for use may be divided into different subsets based on the types of cellular antigens recognized. As an illustration, antigens found on the major leukocyte subtypes including T cells, B cells, antigen-presenting cells, NK cells, and granulocytes, as well as relevant receptors and structures found on these cells are included. These may include activation molecules, co-stimulatory molecules, adhesion molecules, antigen receptors, cytokine receptors, etc. A representative, but not exhaustive, list of the antigens that may be evaluated for RA is provided in Table 1.

Cellular Assay Formats

The cellular assays described above are designed in either of two formats, whole blood or RBC-lysed blood. In the preferred embodiments, the assays are done in whole-blood or RBC-lysed blood format. The minimal manipulation ensures that the most accurate absolute cell counts (cells/μl of blood) are obtained. Furthermore only small amounts of blood are required per assay so that many assays can be run from a single tube of blood. However, for some cell populations an alternative assay format, RBC-lysed blood, will be preferable. These include particular antigen-antibody pairs for which soluble factors (free Ig, soluble cytokine receptors, etc.) contained in the sera interfere with cell labeling and populations of cells that are present in very low frequency. This procedure is useful for activated cells expressing CD25 or CD69 which are essentially undetectable in whole blood from normal individuals but are increased ten-fold in the lysed format and have been shown to be increased in various autoimmune states. Improved detection of other minor cell populations such as NK cells has also been demonstrated and should prove particularly useful in analyses. As an example, for a panel of 96 assays, it is estimated that 64 will be done on whole blood and 32 on lysed blood. Alternative sample processing may include, preparation of PBMC by Ficoll gradient, ex vivo stimulation with polyclonal or antigen specific activators.

Combining antibody reagents is important for the identification of novel cell populations that may contribute to the pathogenesis, or be a marker for, diseases or medical conditions, such as autoimmune diseases. For example, it is known that adhesion molecules can be differentially expressed on T cells thought to be involved in the autoimmune process. Furthermore, several studies have indicated that there may be an increase in the number of memory CD4 T cells in patients with autoimmune disease. With the assays of the present invention it is possible to simultaneously look at differential levels of adhesion molecules (e.g., $CD11a^+$) specifically on a subset of memory (i.e. $CD45RO^+$) T cells of the HLA class II-restricted lineage (i.e. $CD4^+$). This should increase the ability to identify relevant disease-related cell populations. Multiple-color capability also allows one to look for novel populations of cells by choosing combinations of antigens not typically found together on a given cell type or markers found on the same cell type at different stages of ontogeny.

Determination of Appropriate Fluorescent Dyes

As indicated above any appropriate fluorescent dye is within the scope of the present invention. Two commonly used dyes are cyanine dyes Cy5 (em 667) and Cy5.5 (em 703). Typically, a single dichroic filter to split the emission signal at 685 nm is used. More filters will be required when more than two dyes are employed. Dyes are evaluated to determine their compatibility in the MLSC system. As an example, a variety of dyes were evaluated to determine an appropriate overall 3-color set (see table 2). Parameters to consider when evaluating dyes include 1) spectral separation of the 3 dyes, 2) signal-to-noise ratio as a function of laser power, 3) suitability of the available filters, 4) ease of conjugation, and 5) specificity of the resulting antibody-fluorophore conjugates. Cy5 and APC are appropriate for the first color and Cy5.5 is appropriate for the second color. Several potential dyes are appropriate for the third color. Cy7-APC is expected to be suitable for the MLSC system. Preliminary results with the Imagn 2000 system demonstrate that this dye is detectable in the long wavelength channel (>685 nm) and distinct from both Cy5 and APC. Emission spectra indicate that overlap with Cy5.5 should not be a problem given appropriate filters for the new instrument. Fluorescent microspheres offer a wide variety of alternative colors and have been used successfully in some cytometry applications. Conjugation methods will be used which minimize the non-specific binding that occasionally occurs with microsphere reagents. Typically, each of the

fluorophores are evaluated in the context of fluorophore-antibody conjugates using a few select antibodies e.g. anti-CD3, anti-CD4 and anti-CD20.

Soluble Factor Approaches

There are a large number of bioanalyses carried out by means other than fluorescence. Prominent among these is mass spectrometry, rapidly becoming the tool of choice for detailed identification and analysis of polypeptides and proteins. There are two widely-used methods for biomolecular sample introduction in mass spectrometry: electrospray ionization(ESI) and matrix-assisted laser desorption/ionization (MALDI). MALDI-TOF is currently successfully utilized for the analysis of proteins, polypeptides and other macromolecules. Even though the introduction of an organic matrix to transfer energy to the analyte has advanced tremendously the field of desorption mass spectrometry, MALDI-TOF still has some limits. For instance, the detection of small molecules is not practical because of the presence of background ions from the matrix. In such cases ESI or even gas chromatography (GC) mass spectrometry can be used to detect or profile.

The complexity of molecular structures and heterogeneous nature of proteins necessitates the need for multidimensional separation techniques. One of the major areas for this is the development of two dimensional gel electrophoresis using polyacrylamide as the gel matrix for example. The gel is modified in terms of crosslinking, addition of detergents, immobilization of enzymes or antibodies (affinity electrophoresis) or substrates (zymography) and pH gradient. This technique is used for the characterization of proteins in terms of structural modifications, activities, pI values, and molecular weights.

Another area is the development of multidimensional chromatographic approaches also referred to as hyphenated separation techniques. The advantages include the ability to more accurately quantify the analyte and better compatibility with online detection methods like laser induced fluorescence or mass spectrometry. To date usually two separation systems are chosen such that they are orthogonal and lead to a better peak capacity (resolution). Major technical hurdles are the integration of the various separation techniques with the detection system in terms of maintaining resolution upon transfer to the second dimension and the compatibility of the mobile phases with the detection system, for examples salts and detergents in the eluant are incompatible with electrospray mass

spectrometry. Naylor J. Chromatogr. A 1996, 744 237-78 ; Jorgenson Anal. Chem 1997, 69, 1518-1524.

Chemical derivatization can be selectively employed to activate components in a mixture that are not ionized enough to yield an ESI mass spectra. For example sterols typically devoid of acidic or basic residues that do not ionize under electrospray conditions have been coupled with ferrocene carboxylic acids, the electrochemical nature facilitating ionization. Anal. Chem. 1994, 66, 209-212. Derivatization can also serve as an handle to differentiate between stereoisomers (isobaric species) by using different fragmentation patterns in their daughter and granddaughter ions of the parents. The MSn capability in an quadrupole ion trap mass spectrometer for example has been used to distinguish hexosamine monosaccharides, glucosamine, galactosamine and mannosamine derivatized with CoCl2(DAP)2Cl where DAP is diaminopropane. Anal. Chem 1999, 71, 4142-4147.

Affinity based separation followed by mass spectrometric detection is of clinical interest as it allows analysis of complex molecules in biological fluids like blood and urine with little or no sample preparation. Ciphergen's technology (surface enriched laser desorption ionization (SELDI), a variation of MALDI) is based on this principle. Ciphergen offers 5-6 different surfaces upon which protein and/or small molecules are applied, and then washed with increasing stringency. Since each surface/stringency combination leads to a different adsorption profile, the technique provides means for analysis of a complex mixture.

Clinical Data and Informatics

The identification and correlation of biological markers with clinical measurements requires the integration of vast amounts of biological and medical data and a search engine that makes such data accessible and usable. The instrumentation and assays developed in the present invention have the ability to identify hundreds to thousands of independent markers from a small sample of blood. The present invention includes developing a broad clinical strategy to collect extensive medical information from patients that are followed over the time of disease progression and response to therapy. In addition, the present invention includes software, databases and data mining tools to correlate patterns of markers with specific diseases, disease progression and responses to therapy, including, but not limited to, databases of assays and clinical information, data conversion and statistical analysis tools, and medical questionnaire prototypes. The information system of the

present invention is designed to use common language and common formats for entry of disparate types of data and is structured for data-mining purposes.

The universal medial language which will likely become widely used in the next several years is SnoMed-RT. This language will be readily adaptable with the current information system of the present invention. Similarly, the present invention is adaptable in that as other languages or technologies become available, they may also become incorporated into the database. An example would be the eventual development of tools to integrate digital x-rays, mammograms, or a virtual colonoscopy which is obtained via a C.T. scan of the abdomen.

The technical challenges in developing an informatics system capable of handling the vast amounts of biological and clinical information necessary to correlate biological markers with disease include modeling and integrating a number of diverse, complex, and often incompatible information sources, adapting to rapid advances in scientific and medical knowledge and methods, and developing a user-friendly interface, proper format and powerful search tools. The informatics system provided by the present invention meets these technical challenges.

Data Analysis

The data output from the cellular analyses includes both numbers of cells per μl of whole blood for each population identified, the mean intensity of staining for each cell associated molecule, which gives an estimate of the antigen density for a given population, the mean size of cells, and the expression levels of a particular molecule. Each number will be analyzed, because, as explained above, both the actual cell numbers as well as expression levels of a particular molecule may vary in a given disease state. To identify markers (cell counts or staining intensity, levels of soluble factors) associated with categorical clinical variables (such as diagnosis of disease) a variety of discriminant techniques are used. To identify markers associated with continuous-valued clinical variables (such as levels of soluble factors) a variety of regression techniques are used. For both discriminant and regression analyses, stepwise variable selection and cross-validation are used to identify those markers that are most closely associated with the clinical variable of interest. Where appropriate, demographic and clinical variables (such as age, gender, concomitant drugs, etc.) and genetic parameters are included as covariates in the models.

These techniques are applied in the analyses of data, typically, using the SAS and Statistica statistical analysis software packages.

The architecture of the integrated informatics infrastructure (see Figure 3) of the present invention, comprises a multi-tiered structure. The lowest level consists of a set of data sources. The first source comprises the scientific data which includes, but is not limited to, cellular assay data and soluble factor assay data. The second source may be semi structured data which is in a combined form of textual and tabular data describing protocols for assay development and protocols for the execution of clinical studies. The structure may be encoded as a data type definition (DTD), defining tags that serve both for information indexing and querying as well as selective information display on web browsers. The DTD tags also define an information exchange model enabling the high-level electronic sharing of the information with other parties. The third data source is the clinical data gathered and restructured to meet the clinical study requirements. Clinical questionnaires that are optimized to maximize, under time constraints, the collection of useful and quantifiable information from patients, are used to gather information and to provide the necessary quality control. If necessary, the questionnaires will be multi lingual and adapted to physical challenges (e.g., the inability to use a computer keyboard) that the respondents may have to face. The technology of choice for this data source may be XML. In addition, the clinical information gathering system also comprises of non-textual means of input. A respondent may interact via visual and graphical displays to provide health related information by pointing at images of the human anatomy so as to indicate a problem without having to articulate it. Other means, e.g. a simple measure of vital capacity and FeV1sec. in asthmatic or emphysematous patients or monitoring devices to detect and/or correct cardiac arrhythmias, etc. could also be used for input. The fourth source of data is the instrumentation data containing all of the relevant parameter settings required for the execution of the scientific assays on a combination of different instruments, such as Imagn, SurroScan and the ELISA plate reader.

Additionally, data can be collected and recorded in lists. In list form, measurement values for each individual cell are recorded. This facilitates identification and analysis of individual cell populations that express a complex set of different molecules. Alternative analysis schemes are readily explored, facilitating optimal data analysis. Likewise, the complete set of patient data (cell populations, soluble factors, medical history, clinical parameters, etc.) can be stored in lists for each patient sample.

As indicated in Figure 3, these data sources are integrated and warehoused using a common schema. This schema coordinates the interpretation of the information from the constituent data sources. The interpretation is in a manner that is independent of the logical or physical storage detail of each of the constituent data sources. The common schema provides that data sources can be added or modified over time (management of change) without significantly affecting the tool set or user interface that ultimately use the compiled data. The common schema provides a buffer between the ever changing data sources and the application programs which use the compiled data and derive knowledge from the data. Similarly, if in the future additional instrumentation, e.g., NMR, is included for the generation of additional data, it can be added without upsetting the organization of the already existing warehouse.

The schema is augmented with an ontology of common concepts and their relationships in immunology and related clinical areas. The ontology will be used by the data mining tools and by the user interface to assist in the interpretation of user specified requests for information from the underlying data sources and for the specification of data mining tasks. The ontology will also be utilized in the verification of the collected clinical data.

The toolkit of programs includes programs for statistical analysis, for data mining and for the visualization of the results. A result of the analysis by the toolkit programs provides a set of rules relating a set of conditions to a set of consequences. These rules are applied over a statistically significant portion of the underlying data and are of the form:

if cond1 and cond2 and ... and condN then consequence1, consequence2, ...

For example, when applied to the cellular data source and the clinical data source, the toolkit can derive relationships between cellular assay and soluble factor measurements that were previously unknown. The results of the analysis by the toolkit are recycled to the users and to the database reuse in the future. The architecture is intended to improve the knowledge discovery process by storing the accumulated discovery experience and by integrating this experience for continued improvement.

Other tools for data mining include methods for clustering in highly dimensional data. These tools are intended to augment and replace the existing method of manual gating as presently used. Unlike current cytometry software, which considers one assay from one patient at a time, the cytometry tools of the present invention may examine list mode data across an assay from multiple patient samples in order to determine the optimal set of

circumscribed population (gates). The system is coupled with a multi dimensional visualization system that will simultaneously project the computed clusters on selected subsets of two-dimensional and three-dimensional views.

The final tier is a user interface. This part of the system serves the user interaction and is used to plan and execute tasks related to clinical studies. Tasks supported at the user interface level include, knowledge discovery from study data, clinical study planning, protocol planning and evaluation and assay development.

The user interface will accept requests for information in a uniform way. It may combine a graphical interface and may allow for "drilling down" of information from the abstract concept level to the stored detail. It may allow for information requests that include both data and text (e.g., documentation pertaining to assay protocol planning) and may allow for interaction over a network.

EXAMPLES

Example One

Use of the present invention to identify biological markers for Rheumatoid Arthritis

The present invention can be used to identify biological markers for rheumatoid arthritis (RA). Microvolume laser scanning cytometry (MLSC) is used to help create data for identifying biological markers for RA. Marker discovery efforts are focused on readily accessible biological fluids, most notably blood. A two-color instrument and antibody-based assays have demonstrated the potential of this technique for identifying and enumerating scores of different cell populations with only a small amount of whole blood. Multiparameter cell analysis, in combination with multiple assays for soluble factors, small molecules and an extensive clinical database, is a powerful tool for future biological marker discovery. Such markers have the potential to lead to new and more effective ways to predict and monitor disease activity and responses to therapy.

Rheumatoid Arthritis is a chronic inflammatory disorder of the small joints, which also has pronounced systemic consequences. Although the etiology of the disease is unknown, its pathology evolves with common characteristics over time. Early events appear to include an inflammatory response initiated by unknown mediators. Activated $CD4^+$ T cells appear to amplify and perpetuate the inflammation. The presence of activated T cells can induce polyclonal B-cell activation and production of Rheumatoid Factor (RF). Tissue damage accrues, releasing autoantigens, and the extent of the T cell

response broadens. Eventually, the constant inflammatory environment may lead to transformation of the synovial fibroblasts, yielding destructive potential that is independent of T cells and macrophages. The pro-inflammatory cytokines, produced mainly by macrophages in the joint and the cytokines they induce such as IL-6, are systemically active, present in the serum and augment hepatic synthesis of acute-phase proteins. Throughout the various stages of the disease, there are changes in the molecules and cells in the synovium and blood that have potential to be markers of disease. Blood, because of its ready accessibility and circulation throughout the body, provides an attractive window for monitoring disease activity and is thus the major target of this invention.

The present invention is useful to identify biological markers of diagnostic and prognostic value for Rheumatoid Arthritis. Such markers are required for classifying different forms of the disease, for example identifying the subset of patients in whom joint erosion occurs more rapidly than in others. Furthermore, the markers are critical for evaluating the efficacy of intervention and developing early, non-toxic and successful therapies. Many investigations have been made of cells and soluble factors in blood, synovium and urine that are candidate markers for the disease. In general, one to several markers at a time have been investigated. While some factors, such as rheumatoid factor and C-reactive protein have been associated with RA, there is no consensus panel of RA-specific markers. There is a strong need to simultaneously evaluate multiple candidate markers. This is achieved with multiple assays and by increasing the number of parameters (colors) that can be measured in a single assay.

The present invention is capable of developing a platform for identifying markers of disease and applying it to RA. In general, each assay combination consists of one or more reagents to identify the major cellular subsets (left column of Table 1). Some of these antigens, e.g. CD4, are targeted in multiple assays. The major markers are combined with different subsetting antibodies (right column of Table 1) in order to maximize information about the sample. Properties of the fluorochomes and the target antigens are considered in developing each assay combination. For example, brighter fluorochromes are used with less abundant antigens. For other assays it is important to use reagents with the best spectral differences for certain targets. In general for each antibody triplet FlowJo software is used to analyze 1 to 3 different 3-color combinations (e.g., Cy5 CD3, Cy5.5 CD4, Cy7APC-CD45RA vs. Cy5 CD45RA, Cy5.5 CD4, Cy7APC-CD3) to determine the best combination for distinguishing the different cell populations.

Designing a successful panel of assays requires some empirical knowledge. The process is typically an iterative one, with each experiment building upon the previous one. As an example, an overview of candidate combinations with potential value for RA is given below.

T cells. The major antigens being evaluated in a T cell panel include CD2, CD3, CD4, CD5, CD7, and CD8. Many kinds of molecules on these T cell subpopulations can be investigated. These include surface antigens which help to distinguish naïve (CD45RA) vs. memory cells (CD45RO, CD26), and antigens that play a role in activation (CD25, CD69, CD71, HLA class II) or co-stimulation (CD27, CD28). In addition, markers that may play a role in adhesion to inflammatory sites are assayed (CD62L, CD11a/CD18, CD44, CD54, and CD58). Subpopulations of T cells based on expression of αβTCR, γδTCR, and a panel of Vβ TCR genes are evaluated.

B cells. The major antigens being evaluated in a B cell panel include CD19, CD20, CD21, CD22, CD23, and CD72. In addition, various markers on these B cell subsets including markers that may indicate a more activated phenotype (CD40, CD80, CD86, HLA class II, CD5) and those that have been implicated in lymphocyte homing and adhesion (CD62L, CD44, CD11a/CD18) are analyzed. IgM, IgG, and IgA receptors for specific antigens are also evaluated.

Antigen-presenting cells. Antigen-presenting cells are evaluated using markers to the major antigens CD13, CD14, CD15, and CD33. In addition, a variety of adhesion molecules (CD11a, CD18, CD29, CD44, CD54, CD58, CD62L) and co-stimulatory molecules (CD80, CD86) on these cells are analyzed. Other relevant receptors including CD16 (FcγRIII), CD32 (FcγRII), and CD64 (FcγRI) are assayed.

Other cell types and antigens. Only a few studies have investigated the expression of NK markers and granulocyte markers in RA, and in general these have given inconsistent results. NK subpopulations using the markers CD16, CD56, CD57, and NKB1 are analyzed. Granulocytes, including neutrophils and eosinophils, may be phenotyped using CD13, CD15, and CD16. A panel of adhesion molecules and receptors similar to that described above is used to further subset these populations.

There are many antigens whose expression has been associated with a more activated or memory phenotype, implicated in adhesion or co-stimulation, or shown to be the receptor for an important ligand. Examples are outlined in Table 1. Several of these markers have been examined in several autoimmune conditions and the expression has

been found to be variable. For example, T cells from RA patients show higher levels of the adhesion receptor LFA-1 (CD11a/CD18) but no change in the expression of the IL2 receptor (CD25), which is normally increased on activated cells, or a marker for activation and co-stimulation (CD80).

Some examples illustrating the kinds of indicators and cell populations that may be examined are discussed below.

T cells. There are several lines of evidence that implicate T cells in RA (Fox, D.A. (1997) Arthritis Rheum 40, 598-609). Such evidence includes the association of RA with MHC class II alleles that share a common sequence in the third hypervariable region (Weyand, C.M. and Goronzy, J.J. (1997) Ann N Y Acad Sci 815, 353-6 and Weyand, C.M. and Goronzy, J.J. (1997) Med Clin North Am 81, 29-55). Since $CD4^+$ T cells recognize antigen bound to MHC class II antigens, the association of RA with expression of specific class II molecules implies a role for CD4 T cells in RA. In addition, studies in animal models of RA, such as collagen induced arthritis or adjuvant arthritis, have shown that T cells transferred from affected animals can induce synovitis in susceptible hosts. Furthermore, studies in RA patients have shown that strategies aimed at eliminating T cells or interfering with T cell function can ameliorate rheumatoid inflammation.

Perhaps more relevant to the present invention, examination of the phenotype of T cells, either in the synovial fluid, synovial tissue and/or peripheral blood of RA patients, have led to some interesting findings (Cush, J.J. and Lipsky, P.E. (1991) Clin Orthop , 9-22). Increased numbers of activated T cells are detectable in the peripheral blood and synovial fluid of RA patients. These T cells express CD3 and CD4 cell surface markers at a lower antigen density compared to controls, similar to the levels seen in mitogen activated T cells *in vitro* (Luyten, F., Suykens, S., Veys, E.M., Van Lerbeirghe, J., Ackerman, C., Mielants, H. and Verbruggen, G. (1986) J Rheumatol 13, 864-9). There is also a slightly decreased number of $CD8^+$ cells in most active RA causing an increase in the CD4/CD8 ratio. In addition T cells from patients with RA express increased amounts of the early activation marker CD69 (Pitzalis, C., Kingsley, G., Lanchbury, J.S., Murphy, J. and Panayi, G.S. (1987) J Rheumatol 14, 662-6), increased numbers of $CD4^+CD29^+$ and $CD4^+CD45RO^+$ memory cells, and increased expression of MHC class II products (Pitzalis, C., Kingsley, G., Murphy, J. and Panayi, G. (1987) Clin Immunol Immunopathol 45, 252-8). Expression of CD44-dependent primary adhesion strongly correlates with concurrent symptomatic disease in juvenile RA and systemic lupus erythematosus (Estess,

P., DeGrendele, H.C., Pascual, V. and Siegelman, M.H. (1998) J Clin Invest 102, 1173-82) and may be important in adult RA. Some studies have shown increased numbers of γδ TCR T cells and increased HLA expression on these cells (Reme, T., Portier, M., Frayssinoux, F., Combe, B., Miossec, P., Favier, F. and Sany, J. (1990) Arthritis Rheum 33, 485-92). An increase in $CD8^+57^+$ cells in RA, sometimes associated with restricted TCR, has also been reported (Morley, J.K., Batliwalla, F.M., Hingorani, R. and Gregersen, P.K. (1995) J Immunol 154, 6182-90 and Serrano, D., Monteiro, J., Allen, S.L., Kolitz, J., Schulman, P., Lichtman, S.M., Buchbinder, A., Vinciguerra, V.P., Chiorazzi, N. and Gregersen, P.K. (1997) J Immunol 158, 1482-9). Three-color assays can monitor restricted Vβ expression on this specific T cell subset.

B cells. Phenotypic analysis of B cells has also been performed in RA patients. A B cell subpopulation expressing the pan T cell marker CD5 has been shown to be elevated (Sowden, J.A., Roberts-Thomson, P.J. and Zola, H. (1987) Rheumatol Int 7, 255-9, Hardy, R.R., Hayakawa, K., Shimizu, M., Yamasaki, K. and Kishimoto, T. (1987) Science 236, 81-3 and Casali, P., Burastero, S.E., Nakamura, M., Inghirami, G. and Notkins, A.L. (1987) Science 236, 77-81). This subset is also elevated in autoimmune mice where IgM autoantibodies have been shown to be constitutively expressed (Hayakawa, K. and Hardy, R.R. (1988) Annu Rev Immunol 6, 197-218). In humans however $CD5^+$ B cells do not preferentially produce autoantibodies (Suzuki, N., Sakane, T. and Engleman, E.G. (1990) J Clin Invest 85, 238-47) and the role of $CD5^+$ B cells in the pathogenesis of autoimmunity in humans is still unclear, perhaps reflecting the presence of activated B cells (Werner-Favre, C., Vischer, T.L., Wohlwend, D. and Zubler, R.H. (1989) Eur J Immunol 19, 1209-13). Circulating B cells from RA patients also demonstrate increased expression of HLA DR molecules, again indicative of an activated B cell phenotype (Eliaou, J.F., Andary, M., Favier, F., Carayon, P., Poncelet, P., Sany, J., Brochier, J. and Clot, J. (1988) Autoimmunity 1, 217-22). Three-color assay are able to monitor increased HLA class II expression specifically on $CD5^+$ $CD19^+$ B cells.

Antigen-presenting cells. Several cell types can serve as antigen-presenting cells, including monocytes, macrophage, dendritic cells, B cells and other cells induced to express class II antigens. In general these cells show an activated phenotype demonstrated by increased expression levels of HLA class II antigens in patients with autoimmune disease (Lipsky, P.E., Davis, L.S., Cush, J.J. and Oppenheimer-Marks, N. (1989) Springer Semin Immunopathol 11, 123-62). Antigen-presenting cells are abundant in the synovial

compartment (Viner, N.J. (1995) Br Med Bull 51, 359-67) and blood-derived macrophages have been associated with human cartilage glycoprotein 39 expression in some studies (Kirkpatrick, R.B., Emery, J.G., Connor, J.R., Dodds, R., Lysko, P.G. and Rosenberg, M. (1997) Exp Cell Res 237, 46-54).

Soluble factor assays. Soluble factor assays provide an additional battery of potential biological markers. There are many important soluble factors that have been identified in RA patients. These include levels of circulating cytokines such as TNFα and IL-6, cytokine receptors, chemokines, rheumatoid factors of different isotypes, immunoglobulin with different forms of glycosylation, hormones, acute-phase proteins such as C-reactive protein and serum amyloid A, and soluble adhesion molecules, as well as matrix metalloproteinases and their inhibitors. Many of these soluble factors are known to be present at varying levels in RA patients at different stages of disease (Choy, E.H. and Scott, D.L. (1995) Drugs 50, 15-25, Feldmann, M., Brennan, F.M. and Maini, R.N. (1996) Annu Rev Immunol 14, 397-440, and Wollheim, F.A. (1996) Apmis 104, 81-93). Therefore, assays can be conducted to measure these soluble factors and look for statistical correlations with the cell populations identified.

Medical histories. In addition to soluble factors, information in the medical history of patients are included in the database. The clinical parameters will include information on age, gender, stage of disease, outside laboratory tests such as ESR, previous therapy and any concomitant drugs or therapies. This information is relevant to the evaluation. For example, it is known that immunosuppressive drugs, such as those often taken by RA patients, can have a profound effect on the expression of cell surface antigens. Patients treated with methotrexate show a decrease in $CD19^+$ and $CD5^+19^+$ B cells. Patients treated with cyclophosphamide show a decrease in activated T cells expressing CD25 or HLA DR. Patients treated with prednisone express several changes in cell surface phenotype, including a decrease in activated $CD3^+25^+$ T cells, a decrease in $CD5^+19^+$ B cells, and a decrease in $CD16^+$ and $CD56^+$ NK cells (Lacki, J.K. and Mackiewicz, S.H. (1997) Pol Arch Med Wewn 97, 134-43). Other clinical variables such as disease duration may also be useful.

Distinguishing patient populations. A review of the cellular assay literature as it relates to autoimmune disease reveals that there are apparently conflicting reports. For example, some reports indicate an increase in levels of CD5 B cells (Markeljevic, J., Batinic, D., Uzarevic, B., Bozikov, J., Cikes, N., Babic-Naglic, D., Horvat, Z. and Marusic,

M. (1994) J Rheumatol 21, 2225-30) in RA patients, while other studies do not (Liu, S.T., Wang, C.R., Liu, M.F., Li, J.S., Lei, H.Y. and Chuang, C.Y. (1996) Clin Rheumatol 15, 250-3). These publications suggest that there may be other confounding factors that have important implications for cellular phenotypes, and perhaps cellular function, in RA patients. Segregating the patient populations selected for study, based on levels of soluble factors circulating in the serum, stage of disease, and therapy could in part explain the apparent discrepancy with respect to the $CD5^{+}19^{+}$ B cell levels in RA patients discussed above. For example, it is known that there is a significant correlation between the levels of IgM rheumatoid factor and the percentage of CD5 B cells (Youinou, P., Mackenzie, L., Katsikis, P., Merdrignac, G., Isenberg, D.A., Tuaillon, N., Lamour, A., Le Goff, P., Jouquan, J., Drogou, A. and et al. (1990) Arthritis Rheum 33, 339-48). Furthermore the level of IgA rheumatoid factor is associated with the level of CD5 B cells as well as $CD4^{+}CD45RO^{+}$ T cells (Arinbjarnarson, S., Jonsson, T., Steinsson, K., Sigfusson, A., Jonsson, H., Geirsson, A., Thorsteinsson, J. and Valdimarsson, H. (1997) J Rheumatol 24, 269-74). Simultaneous measurement of multiple parameters increases the probability of identifying key variables for segregating patient groups.

This generic example illustrates that this invention is uniquely suited for identifying ensembles of biological markers to characterize diseases. The MLSC technology, which requires only a very small sample volume, provides that numerous assays can be completed on a single blood sample and ensures that the maximum amount of biological information can be acquired. The biological marker identification system can accommodate a mixture of assay types, including whole blood and RBC-lysed blood, among others. The assays conducted are considered relevant for the clinical indication and allow a broad survey. Relevant biological markers can be identified using the technology of the present invention.

Example Two

Use of the present invention to identify biological markers for Multiple Sclerosis

The present invention can be used to identify biological markers for Multiple Sclerosis (MS). The biological marker identification system is employed to identify markers for Multiple Sclerosis. MS is an autoimmune inflammatory disease of the central nervous system. MS is characterized clinically by relapsing and remitting episodes of neurologic dysfunction. The etiology of the disease remains unknown, however the

presence of inflammatory cells in the brain, spinal cord, and cerebrospinal fluid implies that an immune attack against CNS myelin is central to the pathogenesis of MS. The hallmark of the MS lesion is an area of demyelination called a plaque that may be found throughout the brain and spinal cord. Inflammatory cells are seen at the edges of the plaque and scattered throughout the white matter. The main inflammatory cells include activated lymphocytes and monocyte derived macrophages. CD4 T cells accumulate at the edges of the plaque; CD8 T cells are not found as frequently in active disease, but are present in longstanding lesions. Autoreactive T cells recognizing myelin basic protein and other non-myelin self-antigens circulate in the blood and upon activation can pass through the blood-brain barrier. Up-regulation of adhesion molecules, histocompatability antigens, and other markers of lymphocyte and monocyte activation (IL2R, FcR) are all connected with the activation and homing process. Furthermore, there is in increase in proinflammatory cytokines that serves to amplify the immune response. The autoimmune response also includes pronounced B cell stimulation. The autoantibodies produced can activate the complement system and promote demyelination. Throughout the various stages of disease, there are changes in the molecules and cells in the CNS and the blood that have potential to be markers of disease.

The present invention can identify disease markers of diagnostic and prognostic value for Multiple Sclerosis. Such markers are valuable for classifying different forms of the disease, for example identifying the subset of patients with relapsing-remitting disease who are most likely to develop those secondary progressive disease. Furthermore, the markers are valuable for evaluating the efficacy of intervention and developing early, non-toxic and successful therapies. Many investigations have been made of cells and soluble factors in blood, cerebrospinal fluid (CSF) and urine that are candidate markers for the disease. In general, one to several markers at a time have been investigated. While some factors, such as oligoclonal immunoglobulin in the CSF, have been associated with MS, there is no consensus panel of MS-specific markers. There is a strong need to simultaneously evaluate multiple candidate markers.

<u>T cells</u>. There are several lines of evidence that implicate T cells in MS. Such evidence includes the association of MS with MHC class II (particularly HLA DR) alleles (Hauser, S.L., Fleischnick, E., Weiner, H.L., Marcus, D., Awdeh, Z., Yunis, E.J. and Alper, C.A. (1989) Neurology 39, 275-7). Since $CD4^+$ T cells recognize antigen bound to MHC class II antigens, the association of MS with expression of specific class II molecules

implies a role for CD4 T cells in MS. In addition, studies in animal models of MS such as mouse or rat experimental allergic encephalomyelitis have shown that myelin antigen specific CD4 T cells can induce disease when adoptively transferred to naïve animals (Cross, A.H. and Raine, C.S. (1990) J Neuroimmunol 28, 27-37 and Cross, A.H., Cannella, B., Brosnan, C.F. and Raine, C.S. (1990) Lab Invest 63, 162-70). Furthermore, studies in MS patients have shown that strategies aimed at eliminating T cells or interfering with T cell function can slow progression of MS.

Perhaps more relevant to the present invention, studies examining the phenotype of T cells, either in the cerebrospinal fluid and/or peripheral blood of MS patients have led to some interesting findings. There is a reduction in $CD8^+$ T cells in the blood of MS patients. The subset showing the most marked decrease was the $CD8^+CD11b^+$ subset (Ilonen, J., Surcel, H.M., Jagerroos, H., Nurmi, T. and Reunanen, M. (1990) Acta Neurol Scand 81, 128-30 and Oksaranta, O., Tarvonen, S., Ilonen, J., Poikonen, K., Reunanen, M., Panelius, M. and Salonen, R. (1996) Neurology 47, 1542-5). There is also an increase in activated T cells bearing the CD71 and CD25 markers particularly in active MS (Genc, K., Dona, D.L. and Reder, A.T. (1997) J Clin Invest 99, 2664-71 and Strauss, K., Hulstaert, F., Deneys, V., Mazzon, A.M., Hannet, I., De Bruyere, M., Reichert, T. and Sindic, C.J. (1995) J Neuroimmunol 63, 133-42). Furthermore, the majority of T cells in the cerebrospinal fluid and peripheral blood show a memory phenotype with high levels of CD45RO and CD29 on both the CD4 and CD8 T cell populations (Vrethem, M., Dahle, C., Ekerfelt, C., Forsberg, P., Danielsson, O. and Ernerudh, J. (1998) Acta Neurol Scand 97, 215-20). This leads to a reduction in $CD4^+CD45RA^+$ (Strauss, K., Hulstaert, F., Deneys, V., Mazzon, A.M., Hannet, I., De Bruyere, M., Reichert, T. and Sindic, C.J. (1995) J Neuroimmunol 63, 133-42) and $CD8^+CD27\text{-}CD45RA^+$ (Hintzen, R.Q., Fiszer, U., Fredrikson, S., Rep, M., Polman, C.H., van Lier, R.A. and Link, H. (1995) J Neuroimmunol 56, 99-105) naïve T cells in the peripheral circulation. A recent study has concluded that $CD4^+$, $CD4^+$ $SLAM^+$, and $CD4^+CD7^+$ cells (preferentially T helper 1 cytokine producing cells) are increased in MS patients relative to controls (Ferrante, P., Fusi, M.L., Saresella, M., Caputo, D., Biasin, M., Trabattoni, D., Salvaggio, A., Clerici, E., de Vries, J.E., Aversa, G., Cazzullo, C.L. and Clerici, M. (1998) J Immunol 160, 1514-21). Furthermore some studies have shown skewed TCR variable beta usage in the peripheral blood of MS patients indicative of a restricted TCR repertoire (Gran, B., Gestri, D., Sottini, A., Quiros Roldan, E., Bettinardi, A., Signorini, S., Primi, D., Ballerini, C., Taiuti, R., Amaducci, L. and Massacesi, L. (1998)

J Neuroimmunol 85, 22-32). A restricted pattern of gene rearrangement has also been described in the $\gamma\delta$ T cell subset (Michalowska-Wender, G., Nowak, J. and Wender, M. (1998) Folia Neuropathol 36, 1-5).

B cells. Phenotypic analysis of B cells has also been performed in MS patients. A B cell subpopulation expressing the pan T cell marker CD5 has been shown to be elevated (Mix, E., Olsson, T., Correale, J., Baig, S., Kostulas, V., Olsson, O. and Link, H. (1990) Clin Exp Immunol 79, 21-7). This subset is also elevated in autoimmune mice where they have been shown to constitutively express IgM autoantibodies (Hardy, R.R., Hayakawa, K., Shimizu, M., Yamasaki, K. and Kishimoto, T. (1987) Science 236, 81-3). In humans, however, $CD5^+$ B cells do not preferentially produce autobodies (Suzuki, N., Sakane, T. and Engleman, E.G. (1990) J Clin Invest 85, 238-47) and the role of $CD5^+$ B cells in the pathogenesis of autoimmunity in humans is still unclear, perhaps reflecting the presence of activated B cells (Werner-Favre, C., Vischer, T.L., Wohlwend, D. and Zubler, R.H. (1989) Eur J Immunol 19, 1209-13). Consistent with this conclusion, high levels of the memory marker CD45RO were found on circulating $CD20^+$ B cells from patients with MS (Yacyshyn, B., Meddings, J., Sadowski, D. and Bowen-Yacyshyn, M.B. (1996) Dig Dis Sci 41, 2493-8). The number of circulating $CD80^+$ B cells is also increased significantly in MS patients with active disease, but is normal in stable MS (Genc, K., Dona, D.L. and Reder, A.T. (1997) J Clin Invest 99, 2664-71).

Antigen-presenting cells. Several cell types can serve as antigen-presenting cells, including monocytes, macrophage, dendritic cells, B cells and other cells induced to express class II antigens. In general these cells show an activated phenotype demonstrated by increased expression levels of HLA class II antigens in patients with active MS (Genc, K., Dona, D.L. and Reder, A.T. (1997) J Clin Invest 99, 2664-71). A recent study has also shown that CD86 and CD95 (fas) expressing monocytes are increased in MS as compared with healthy controls (Genc, K., Dona, D.L. and Reder, A.T. (1997) J Clin Invest 99, 2664-71).

Other cell types. Only a few studies have looked at the expression of NK markers and granulocyte markers in MS. One study shows a decrease in $CD16^+$ NK cells in chronic, progressive MS (Kastrukoff, L.F., Morgan, N.G., Aziz, T.M., Zecchini, D., Berkowitz, J. and Paty, D.W. (1988) J Neuroimmunol 20, 15-23).

Soluble factor assays. Soluble factor assays provide an additional battery of potential biological markers. There are many important soluble factors that have been identified in MS patients. For example, levels of soluble Apo A-1/Fas (Ferrante, P., Fusi, M.L., Saresella, M., Caputo, D., Biasin, M., Trabattoni, D., Salvaggio, A., Clerici, E., de Vries, J.E., Aversa, G., Cazzullo, C.L. and Clerici, M. (1998) J Immunol 160, 1514-21) is augmented in acute MS compared with the levels seen in patients with stable disease or healthy controls. In addition, levels of soluble adhesion molecules such as soluble intracellular adhesion molecule 1 (ICAM-1) (Giovannoni, G., Lai, M., Thorpe, J., Kidd, D., Chamoun, V., Thompson, A.J., Miller, D.H., Feldmann, M. and Thompson, E.J. (1997) Neurology 48, 1557-65) and soluble E –selectin (Giovannoni, G., Thorpe, J.W., Kidd, D., Kendall, B.E., Moseley, I.F., Thompson, A.J., Keir, G., Miller, D.H., Feldmann, M. and Thompson, E.J. (1996) J Neurol Neurosurg Psychiatry 60, 20-6) have been shown to be increased in MS patients at different stages of disease. Proinflamatory cytokines like TNFα and IFNγ, are known to be present at varying levels in MS patients at different stages of disease (Navikas, V. and Link, H. (1996) J Neurosci Res 45, 322-33). Other relevant proteins, such as cytokines and cytokine receptors, chemokines, matrix metalloproteinases and their inhibitors, neopterin, and myelin basic protein, have also been shown to be present at varying levels in MS patients at different stages of disease and healthy controls. Therefore, assays can be conducted to measure these soluble factors and look for statistical correlations with the cell populations identified.

Medical histories and distinguishing patient populations. In addition to soluble factors information in the medical history of patients will be included in the database. The clinical history will include information on age, gender, stage of disease, outside laboratory evidence (magnetic resonance imaging, cerebrospinal fluid analysis for oligoclonal immunoglobulin and evoked potential recordings), previous therapy and any concomitant drugs or therapies. This information is relevant for segregating patient populations.

It is evident that treatment effects play a role in the phenotype of the cells. While untreated MS patients display a greater population of $CD3^{+}CD4^{+}CD8^{+}$ circulating T cells compared with healthy donors, this population of cells is reduced following corticosteroid treatment [30]. In addition, the number of CD71+ and HLA DR+ lymphocytes and monocytes is increased in active MS. However therapy with IFNβ-1b reduces the number of activated HLADR+, CD71+ and CD25+ cells. Furthermore, although the number of circulating CD80+ B cells is decreased, the number of CD86+ monocytes is increased

following therapy [14]. Other clinical variables such as disease duration may also be useful. For example it has been shown that in MS patients with restricted TCR Vβ repertoires, the median disease duration is shorter than in patients who do not have a restricted repertoire (Gran, B., Gestri, D., Sottini, A., Quiros Roldan, E., Bettinardi, A., Signorini, S., Primi, D., Ballerini, C., Taiuti, R., Amaducci, L. and Massacesi, L. (1998) J Neuroimmunol 85, 22-32).

Example Three

Study Comparing Rheumatoid Arthritis Patient to Control Patient Blood

In a pilot study, a panel of 40 two-color cellular assays was prepared for the Imagn 2000 and evaluated blood samples from about 50 donors. Half of the samples came from the Stanford Blood Bank and half came from the Rheumatology Clinic at Stanford University. The study was designed to evaluate and develop key components of the biomarker search engine of the invention: instruments, assays and analytical tools. It was not necessarily designed to elucidate biomarkers. All assays were done on whole blood, without washing to remove unbound reagents. Thirty-eight of the assays comprised 27 different antibody reagents to 23 different cell surface antigens. Eighteen were conjugated to Cy5 and nine were conjugated to Cy5.5. Each of these cellular assay comprises one antigen conjugated to each dye to make up a two-color combination. Two assays monitored cell viability with a DNA intercalating dye. The cellular assays allowed us to identify approximately 100 different cell populations including sets of T cells, B cells, NK cells, monocytes, and granulocytes.

Methods

Cellular assays

The panel of assays is shown in Table 3. Each reagent is tested and titrated before preparing the reagent combinations in order to optimize assay performance.

Sample preparation

For this study all cellular assays were applied to whole blood, in homogeneous mode (no post stain washing). Aliquots (20 uL) of fluorescently labeled antibody reagent combinations of DNA dye were distributed with a multi-channel pipette from prepared racks into discrete wells of a microtiter plate. Whole blood or diluted whole blood (30 uL) was added with a multi-channel pipette and the sample mixed. Cells were incubated for 20 minutes followed by the addition of 100 uL of diluent and mixing. A portion of each

stained sample (50 uL, corresponding to 10 uL of blood) was added to volumetric capillaries (VC120) and loaded into the modified Imagn 2000 instrument. Scans were initiated and executed without operator intervention. Data files are transferred over the computer network and converted to the Flow Cytometry Standard format. FlowJo cytometry software was used to identify cell populations and obtain numerical values for cell counts (cells per uL), relative cell size, and relative fluorescence intensity, which is an estimate of the antigen density for each gated (boxed) cell population.

Soluble factors

Serum levels of C-reactive protein were measured on the Imagn 2000 with a bead-based immunoassay. Beads coated with anti-CRP antibody were used to capture the analyte. Cy5 conjugated anti-CRP antibody was used reveal the captured analyte.

Patient medical information

An abbreviated medical history, including age, gender, parameters of disease severity, co-morbidities and medications was obtained from each patient. Data for the blood bank samples was limited to age and gender.

Database and statistics

Data output from the cellular assays, soluble factor assay and medical histories were combined into a single database. To identify potential biological markers (cell counts or staining intensity, serum concentration) associated with categorical clinical variables (such as diagnosis of disease) a variety of discriminant techniques was used including Fisher linear and quadratic discriminant analysis, logistic regression, and classification trees. To identify markers associated with continuous-valued clinical variables (such as erythrocyte sedimentation rate) we use a variety of regression techniques including multivariate linear regression and regression trees. For both discriminant and regression analyses, stepwise variable selection and cross-validation was used to identify those markers that are most closely associated with the clinical variable of interest. Where appropriate, demographic and clinical variables (such as age, gender, and concomitant drugs) were included as covariates in the models. These techniques were implemented and applied using the SAS and Statistica statistical analysis software packages.

Results

Common analysis strategies can be used for most patient samples

Although one of the drawbacks of cell population studies is often the variability among donor samples, identical analysis windows (gates) were used across all donors for

95% of the cell populations analyzed. This demonstrates the robustness and consistency of these assays and cell analysis systems. The remaining 5% of the gates were adjusted to account for new populations appearing in certain donors or a reagent that appeared unreliable for a few donors. In the latter case the problem reagent can be replaced with an improved version in future studies.

An example of a 2-color combination with variation among donors is shown in Figure 4. The cells were stained with CD27 conjugated to Cy5 in combination with CD8 conjugated to Cy5.5. This combination allowed $CD8^+$ T cells (MHC class I restricted) to be monitored, which are $CD27^+$ (activated) and $CD27^-$. $CD8^-$, $CD27^+$ cells (which are actually activated CD4, MHC class II restricted, T cells) are also detected. Although there is variation among the donors, a single gating strategy can be implemented. Three cell populations are identified which differ among the donors. In Figure 4A the majority of $CD8^+$ T cells are CD27 negative. In Figure 4B the majority of $CD8^+$ cells are CD27 positive. Finally, in Figure 4C, the CD8 population is split between those that are CD27 positive and those that are CD27 negative. FlowJo, our cytometry software, calculates the cell count for each of the gated populations. In addition, the mean fluorescence intensity for each antigen was obtained. This is indicative of the antigen density on the cell surface. The relative cell size for each cell population was also obtained. The numbers were compared and compute statistics across all donors. The differences shown here with respect to CD27 expression on $CD8^+$ cells are typical of the kinds of changes that are observed when comparing patient and control populations in our clinical study.

Excellent correlation among related measurements

Another goal of this initial study was to assess the robustness of the 2-color Imagn system and develop statistical tools. The study was designed so that several capillaries contained the same antibody reagent conjugated to either the same or the alternative dye. This allowed the same measurement to be obtained anywhere from 2-6 times from the same donor for CD3, CD4, CD5, CD7, CD8, CD19, CD20 and CD27 antigens. The same cell populations were also measured using antibodies to different antigens found on them. For example, total T cells were enumerated using CD3 as well as CD5. B cells were enumerated using CD19 as well as CD20, etc. In this preliminary study, excellent consistency was seen both between capillaries containing the identical reagent and capillaries containing different antibodies staining similar cell populations. Correlation coefficients for the same antigen across different capillaries averaged 0.94. The correlation

coefficient was 0.97 for both CD3 vs. CD5 and CD19 vs. CD20. Examples of these correlations are given in Figure 5 and Figure 6.

Differences are observed among RA and blood bank samples

Several measured parameters were used to segregate general blood bank samples and RA patient samples as shown in Table 4. The best single markers accurately segregate 80 to 86% of the sample (7 to 10 incorrect assignments). Some, two cell population pairs segregate 90% of the samples, suggesting that sets of cell populations may be more useful than single cell populations for segregating patient populations.

Example Four

Expanded RA Study

This Example expands the measurement capabilities in an RA study. Cell populations and soluble factors from rheumatoid arthritis (RA) patients were monitored. The RA patients were part of a clinic study, receiving methotrexate and either ARAVA or a placebo. Patients were monitored longitudinally over about 2 months. At each time point, cell population data, soluble factor data, and clinical information was collected.

Cellular assays

Most of the cellular assays are done in whole-blood format as described in Example three. The minimal manipulation ensures that the most accurate absolute cell counts (cells/μl of blood) are obtained. Furthermore only small amounts of blood are required per assay (40 ul) so that many assays can be run from a single tube of blood. However, for some cell populations an alternative assay format, RBC-lysed blood, is preferable. These include particular antigen-antibody pairs for which soluble factors (free Ig, soluble cytokine receptors, etc.) contained in the sera interfere with cell labeling and populations of cells that are present in very low frequency. For example the RBC-lysed sample preparation is useful for activated cells expressing CD25 or CD69 which are essentially undetectable in whole blood from normal individuals but are increased ten-fold in the lysed format and are likely to be increased in various autoimmune states. Improved detection of other minor cell populations such as NK cells has also been demonstrated.

For this protocol, the cellular assays included a panel of 60 2-color combinations comprising 46 whole blood assays and 14 RBC-lysed whole blood. A total of 39 different antibody reagents (30 conjugated to Cy5 and 9 conjugated to Cy5.5), targeting 35 distinct cell surface antigens, were used. All assays are done in homogeneous mode (no wash after

staining). This assay panel enables the identification of more than 150 different cell populations. The reagent combinations and the cell populations that can be identified are provided in Table 5.

Soluble Factor Assays

Sera are aliquoted and frozen for each blood sample for subsequent measurement of multiple soluble factors. These include levels of circulating cytokines such as TNFα and IL-6, cytokine receptors, chemokines, rheumatoid factors (RF) of different isotypes, immunoglobulin, acute-phase proteins such as C-reactive protein and serum amyloid A, and soluble adhesion molecules, as well as matrix metalloproteinases and their inhibitors. The initial panel of 22 soluble factors assayed is shown in Table 6. Additional targets are also provided in Table 6. All assays are done in a sandwich ELISA format using matched antibody pairs to ensure the required sensitivity and specificity.

Patient medical information

A medical history with more detailed disease-specific information is included with each sample in the study.

Example Five

Cellular Assays on a 4-channel MLSC Instument

More assays, with greater information content per assay, can be run on the 4-channel SurroScan instrument. Assays are developed using 3 color reagent combinations. Effective dye combinations include Cy5, Cy5.5 and Cy7 and Cy5, Cy5.5 and Cy7-APC allow simultaneous and independent measurment of three target antigens. Three color combinations facilitate the acquisition of more information per capillary than 2 color combinations by 1) eliminating redundancy (e.g. measuring CD3, CD4 and CD8 in one capillary instead of measuring CD3 + CD4 and CD3 + CD8 in two capillaries) and 2) identifing new populaitons that are defined by the simultaneous expression of 3 antigens (e.g. naïve $CD4^+$ T cells that express both CD45RA and CD62L). Given appropriate fluorescent dyes with distinct emission spectra, it is possible to simultaneously monitor additional target antigens either in the fourth channel, or in some cases, in the existing channels. Figure 7 provides the results of a 3-color assay on the SurroScan instrument.

Assays on the SurroScan instrument can be executed with capillary arrays which use about 1/3 less sample than the VC120 capillaries. For whole blood assays it is possible to process 10 uL or less per 3 color assay, giving the potential for up to 1000

assays per 10 mL tube of blood. For RBC-lysed blood samples with a 10-fold increase in leukocyte concentration, about 100 assays could be done per tube of blood. A panel of 64 3-color assays with 50 or more target antigens is under development using both whole blood and lysed formats. It should allow identification of more than 200 cell populations.

Example Six

Intracellular staining

Intracellular molecules can be measured with MLSC technology. PBMC were cultured for 5 hours in the presence of PHA and ionomycin. Cells were stained with Cy5.5 anti-CD8 to identify cytoxic T cells, fixed, permiablized, and stained with Cy5 anti-inerferon-gamma (IFN-γ) to detect the intracellular cytokine. Data in Figure 8 shows that IFN-γ is detected only in stimulated cells. A control reagent (MOPC) does not label the cells. Among the CD8 T cells, 20 % express intracellular IFN-γ.

Example Seven

Identifying Biological Markers for the Treatment of Allergy and Asthsma

The present invention can be used to identify biological markers for allergic asthma. Asthma is common chronic lung disease of uncertain etiology. It is characterized by inflammation of the airways leading to symptoms of coughing, wheezing, chest tightness, and shortness of breath. These clinical symptoms are thought to be due to hyper-responsiveness of the airways and a long-term inflammatory process causing obstruction of airflow. The disease causes extreme discomfort and can at times be fatal in the absence of appropriate treatment. The clinical manifestations of asthma are thought to result from the superimposition of a variety of environmental factors on genetic predispositions that increase the likelihood of developing asthma. Atopy, the hypersensitivity to environmental allergens, is common in asthma, but not all atopic individuals develop asthma. The relative importance of allergic mechanisms is not completely understood. Corticosteroids (inhaled and systemic) are efficacious in asthma but have associated with perceived and real side effects that limit their usefulness. A more complete understanding of response to corticosteroids might allow for the development of drugs with only local effects within the lungs or drugs that have beneficial effects without side effects.

A study has been designed to identify biological markers of atopy, asthma and the response to corticosteroid therapy. Subjects are screened for four study groups of 20: 1) mild asthmatics who have tested positive to skin test allergens, 2) mild asthmatics who have tested negative to skin test allergens, 3) non-asthmatics who have tested positive to skin test allergens, and 4) non-asthmatics who have tested negative to skin test allergens (healthy subjects). All eligible subjects are entered onto a single-blinded, placebo controlled, randomised parallel study to investigate the effect of the drug prednisone on biological markers after 3 days bid treatment. Blood samples are taken prior to treatment on the morning of day 1 and 12 hours after the last dosing on the morning of day 4.

Subjects undergo rigorous screening including detailed medical history and clinical tests for lung function and allergy. Mild asthmatics have a 1) $FEV_1 \geq 80\%$ predicted, 2) documented diagnosis of asthma or history of any of the following: cough, worse particularly at night, recurrent wheeze, recurrent difficult breathing, recurrent chest tightness and 3) a positive methacholine challenge test (Cockcroft DW, et al Clin Allergy 1977; 7:235 and Juniper EF, et al Thorax 1984; 39:556). Non asthmatics have a 1) $FEV_1 \geq$ 80% predicted 2) no history of asthma and 3) a negative methacholine challenge test. Allergic subjects have a positive skin test to at least one of a panel of allergens.

Examples of clinical data include Haematology: white blood cell count (WBC), red blood cell count (RBC), hemoglobin (Hb), hematocrit (HCT), mean cell volume (MCV), mean cell haemoglobin (MCH), mean cell haemoglobin concentration (MCHC) platelet count, neutrophil count lymphocyte count, monocyte count, eosinophil count, basophil count and ESR - erythrocyte sedimentation rate; blood biochemistry: alkaline phosphatase, alanine transaminase, aspartate transaminase, gamma-glutamyl transpeptidase, albumin, total protein, total bilirubin, urea, creatinine, sodium, potassium, glucose; urinalysis: protein, glucose, ketones, bilirubin, blood, leucocytes; Hepatitis and HIV testing: HIV I and II, Hepatitis B surface antigen, Hepatitis C antibody. All clinical history and test parameters will be included in the master database for statistical analysis, evaluation as covariates and data mining.

Atopic asthma is an immunologic disease mediated by IgE antibodies. Exposure to allergen causes B cells to synthesize IgE, which binds to the high affinity receptor mast cells residing in the mucosa of the airways. On re-exposure to the allergen, antigen-antibody interactions on the surface of the mast cells triggers release of mediators of anaphylaxis stored in mast cell granules, including: histamine, tryptase, PGD_2, leukotriene

C_4 and D_4, and platelet activating factor (PAF). These soluble factors induce contraction of air smooth muscle and cause an immediate fall in the FEV_1. Re-exposure to allergens also leads to the synthesis and release of a variety of cytokines: IL-4, IL-5, GM-CSF, TNF-α, TGF-β, from T cells and mast cells. These cytokines attract and activate B cells, which leads to the production of more IgE, and eosinophils and neutrophils, which produce eosinophil cationic protein (ECP), major basic protein (MBP) and PAF. These factors cause edema, mucus hypersecretion, smooth muscle contraction and increase the bronchial reactivity that is typically associated with the late asthmatic response, indicated by a fall in FEV_1 about 4-6 hours after exposure.

A broad panel of cellular and soluble factor measurements are applied to the subject blood samples with the goal of discovery biomarkers. The study design supplies information of inter-individual variability within groups, and inter-group differences in marker expression. It is believed that the inter-group differences (e.g. allergic non-asthmatic versus non-allergic non-asthmatic) will be greater than inter-individual variability within groups. It is further believed that prednisone therapy will result in significant intra-individual changes in marker expression.

Cellular assays

A panel of 64 three color cellular assay, focusing on immune and inflammatory parameters in the blood, has been prepared and tested for initial atopic asthma. The panel is given in Table 7.

Soluble Factors

The study will also look at a broad panel of soluble factors. Immunoassays, in the sandwich-based chemiluminescent ELISA format, are used for the following targets:

Cytokines, chemokines and their soluble receptors: IL-1 alpha, IL-1 beta, IL-1 RA, IL-1 sRI, IL-1 sRII, IL-2, IL-2sR, IL-3IL-4, IL-5, IL-6, IL-6 sR, IL-8, IL-10, IL-12 p40, IL-12 p70, IL-13, IL-16, IL-17, MIF, MIP-1 alpha, MIP-1 beta, RANTES, sTNFalpha RI *, sTNFalpha RII *, TGF beta, TNF alpha, alpha, TGF beta2, TGF beta3, Oncostatin M, M-CSF, GM-CSF, IGF-1, PDGF-BB, FGF-4, FGF-6, FGF-7, Fas, VEGF, MCP-1, PF-4, EOTAXIN, IFN gamma, Immunoglobulin: IgA1 Kappa, IgA1 Lambda, IgA1, 2 Kappa, IgA1, 2 Lambda, IgA2 Kappa, IgA2 Lambda, IgE total, IgG1 Kappa, IgG1 Lambda, IgG1 total, IgG2 Kappa, IgG2 Lambda, IgG2 total, IgG3 Kappa, IgG3 Lambda, IgG3 total, IgG4 Kappa, IgG4 Lambda, IgG4 total, IgG total , IgG total Kappa, IgG total Lambda, IgM Kappa, IgM Lambda, IgM total, RFIgA, RFIgG, RFIgM, RF total, Acute phase proteins:

CRP, SAA; Matrix metalloproteinases and their inhibitors: MMP-3, MMP-9, TIMP-1, TIMP-2; Soluble adhesion molecules: sCD54 (ICAM-1), sCD62E, sCD62P.

Additional soluble factors which are measured by immunoassays or mass spectroscopy assay include, but are not limited to, Cytokines, chemokines and their soluble receptors: IL-9, IL-11, IL-14, IL-15, IL-18, sCD23, eosinophil proteins: ECP, MBP, Immunoglobulin: Allergen specific IgE, carbohydrate modified Ig; a variety of prostaglandins; a variety of leukotrienes, histamine.

Data output from the cellular assays, soluble factor assays, medical histories and screening labels are combined into a single database. To identify potential biological markers (cell counts, antigen intensity on particular cell types, soluble factor concentrations, etc) associated with categorical clinical variables (disease status, prednisone or placebo, before or after therapy) a variety of ANOVA and discriminant techniques can be used. Where appropriate, demographic and clinical variables (such as age, gender, specific history results) can be included as covariates in the models. Techniques can be implemented with SAS, Statistica, Statview or similar statistical analysis software packages.

<u>Example Eight</u>

<u>Use of the present invention to identify biological markers following the administration of aspirin</u>

The present invention can be used to identify biological markers for evaluating the effects of drug administration on cellular and soluble factors to be performed on small samples of peripheral blood. It is expected that these assays will make possible analysis of the effects of different doses of drugs on cellular and soluble markers in human peripheral blood. In this example the widely used over-the-counter drug, aspirin (acetylsalicylic acid), is administered to human volunteers. Different doses of the drug will be orally administered; blood is drawn before and at various time points after administration, and panels of cellular and soluble factor assays are undertaken. The aspirin is expected to cause changes in the cellular and soluble components of blood.

Aspirin is routinely used for two main indications: 1) to reduce the risk of coronary and cerebral thrombosis and 2) as an analgesic/anti-inflammatory agent. The mechanism underlying the first indication is believed to be irreversible inhibition of the enzyme PGH-synthatase in platelets. A prostaglandin product of this enzyme in platelets is converted to

thromboxane A2, which facilitates platelet aggregation and thrombosis. A side effect of prostaglandin synthesis is the generation of oxygen free radicals, which in the presence of redox-oxidative metals convert unsaturated fatty acids into aldehydes. A relatively stable product of lipid oxidation is malondialdehyde (MDA). This compound is routinely assayed colorimeterically or fluorometrically following interaction with thiobarbituric acid (TBA. Aspirin, by inhibiting prostaglandin synthesis, is expected to decrease MDA levels in peripheral blood platelets. This is one parameter that is expected to change following aspirin administration. Changes in other markers of platelet activation such as changes in the expression of CD62P and CD63 may also occur.

E-type prostaglandins suppress lymphocyte activation and the production of tumor necrosis factor-α (TNF-α) by the cells of the monocyte-macrophage lineage. If there is some level of lymphocyte activation and TNF-α production in normal healthy persons, this may be increased after aspirin treatment and detectable in the peripheral blood. These are examples of expected changes following aspirin administration; if many markers are assayed, unexpected changes may also be found, and may prove to be more interesting than those expected.

The study is designed to identify the effects of aspirin on blood parameters. Eligible subjects are randomly assigned to orally administer aspirin according to one of three dosing schemes. Group I, 1 dose (325 mg tablet) after breakfast, Group II. 2 doses (650 mg) after breakfast and Group III, 2 doses after breakfast and 2 doses after diner (1300 mg total). There are 10-12 subjects per cohort. Blood samples are taken before, during and after aspirin administration. The schedule is given in Table 8. Subjects are healthy individuals age 18-65 who are not taking other aspirin other non-steroidal anti-inflammatory drugs nor currently under care, which requires the use of anti-inflammatory (steroidal or non-steroidal) drugs.

<u>Cellular assays</u>

A panel of 42 three-color cellular assays are used for the initial study see Table 9. The panel includes immune and inflammatory parameters and contains some of the assays listed in Example 7. It also includes a series of assays for platelet function (1-17). These assays include direct measurements in diluted whole blood (WB, 1-9) as well as thrombin stimulation assays (TRT, 10-13) and stimulation controls (NTRT, 14-17).

Soluble Factors

A broad panel of soluble factors as described in Example 7 will be part of the study. Additional measurements include : Von Willebrand factor, b-Thromboglobulin, Thromboxane B2, 6-keto PGF and malondialdehyde. Soluble factors will be measured from plasma. In addition, some soluble factors (e.g MDA, prostaglandins leukotrienes) will be evaluated for the stimulated samples and controls.

TABLE 1

Major Markers	Subsetting antigens
T Cell: CD2, CD3, CD4, CD5, CD7, CD8	**Memory/Activation/C -stimulation:** CD6, CD25, CD26, CD27, CD28, CD38, CD43, CD45RA/RO, CD49a-f, CD69, CD70,CD71, CD80, CD86, CD152(CTLA4), CD154(CD40L)
B Cell: CD19, CD20, CD21, CD22, CD23, CD72	**Adhesion:** CD11a/b/c(integrins), CD18, CD29, CD31, CD44, CD54(ICAM-1), CD58(LFA3), CD62E/L/P(selectins), CD102(ICAM-2), CD104, CD138
Antigen-presenting Cell: CD13, CD14, CD15, CD33	**Antigen receptors:** **TCR:** αβTcR, γδTcR, specific TcR Vβ panel **SIg**: IgM, IgG, IgA
NK cell: CD16, CD56, CD57, NKB1	**FcR:** CD16(FcγRIII), CD32(FcγRII), CD64(FcγRI)
Granulocyte: CD13, CD15, CD16, CD33	**Cytokine receptors:** CD25/CD122(IL2R), CD95(Fas), CD116 (GMCSFR), CDw119(IFNγR), CD120(TNFR), CD121a/b(IL1R), CD123(IL3R), CD124(IL4R), CDw125(IL5R), CD126(IL6R), CD127(IL7R), CDw128(IL8R)
	Nonlineage: CD9, CD35, CD40, CD45, HLA class II DR, DP, DQ, PAN, CDw150(SLAM)

*Some cell surface antigens are in more than one category.

TABLE 2

	Dyes (Excitation/Emission maximum)		
Dye Type	**1st Color**	**2nd Color**	**3rd Color**
Cyanine Dyes[1]	Cy5 (650/667)	Cy5.5 (678/703)	Cy7 (743/770)
Bodipy[2]	BODIPY 630/650-X	BODIPY 650/665-X	
Phycobiliproteins	APC (652/660)		
Tandem Dyes[4]			Cy7-APC (652/780)
PEG Stabilized[5]		La Jolla Blue (680/705)	
Microspheres[2]	Scarlet (645/680)	Dark Red (660/680)	Far Red (690/720) Infrared (715/755) Transfluor (633/720)

[1]Amersham, [2]Molecular Probes, [3] Multiple vendors, [4]PharMingen REF [5] Diatron

TABLE 3

Assay Panel

Combo #	Cy5	Cy5.5	Potential populations detected	Comments
Major T cell subsets (secondarily monocytes based on CD4, NK based on CD7)				
1	CD3/SR0 54	CD4/SR05 1	Total CD4 Total CD3 CD3+4+ (CD4 T) CD3+4- (CD8 T) CD3-4+ (mono)	In Pro-5001, (5 new pops)
2	CD3/SR0 54	CD8/SR12 3	Total CD3 Total CD8 CD3+8+ (CD8 T) CD3+8- (CD4 T) CD3-8+	In Pro-5001, (4 new pops)
3	CD27/SR 162	CD4/SR05 1	Total CD27 Total CD4 CD27+4+ (CD4 T) CD27+4- (CD8 T) CD27-4+ (mono)	New CD27 (4 new pops)
4	CD27/SR 162	CD8/SR12 3	Total CD27 Total CD8 CD27+8+ (CD8 T) CD27+8- (CD4 T) CD27-8+	New CD27 (3 new pops)
5	CD7/SR1 29	CD4/SR05 1	Total CD7 Total CD4 CD7+4+ (CD4 T) CD7+4- (NK + 8) CD7-4+ (mono)	New CD7 (4 new pops)
6	CD7/SR1 29	CD8/SR12 3	Total CD7 Total CD8 CD7+8+ (CD8 T) CD7+8- (NK + 4) CD7-8+	New CD7 (3 new pops)
7	CD5/SR0 52	CD7/SR13 6	Total CD7 Total CD5 CD7+5+ (T) CD7+5- (NK) CD7-5+	New CD7 (4 new pops)
8	αβTCR/S R089	CD7/SR13 6	Total CD7 Total αβTCR CD7+αβTCR+ (T) CD7+αβTCR- (NK) CD7-+αβTCR+	New αβTCR (4 new pops)
Minor T cell subsets (included are CD45RA, CD62L, CD69, CD25)				
9	CD45RA/ SR181	CD4/SR05 1	CD4+45RA- CD4+45RA+	*In Pro-5001* (2 new pops)
10	CD45RA/ SR181	CD8/SR12 3	CD8+45RA- CD8+45RA+	In Pro-5001 (2 new pops)

TABLE 3 (CONTINUED)

11	CD62L/SR098	CD4/SR041	CD4+62L- CD4+62L+.	*In Pro-5001* (2 new pops)
12	CD62L/SR098	CD8/SR123	CD8+62L- CD8+62L+	*In Pro-5001* (2 new pops)
13	CD69/SR099	CD4/SR051	CD4+69- CD4+69+	In Pro-5001 (2 new pops)
14	CD69/SR099	CD8/SR123	CD8+69- CD8+69+	*In Pro-5001* (2 new pops)
15	CD25/SR095	CD3/SR055	CD3+25- CD3+25-	*New CD25* (2 new pops)
B cells and B cell subsets.				
Combo #	**Cy5**	**Cy5.5**	**Potential populations detected by this combination of stains**	**Comments**
16	CD5/SR052	CD19/SR050	Total CD5+ Total CD19+ CD5+19+ (CD5+ B cells)	In Pro-5001 (2 new pops)
17	CD25/SR095	CD19/SR050	CD19+25- CD19+25+	New CD25 (2 new pops)
18	CD69/SR099	CD19/SR050	CD19+69- CD19+69+	In Pro-5001 (2 new pops)
19	CD80/SR101	CD19/SR050	CD19+80- CD19+80+	New CD80 (2 new pops)
20	CD86/SR143	CD19/SR050	CD19+86- CD19+86+	New CD86 (2 new pops)
21	CD62L/SR098	CD20/SR160	Total CD20+ CD20+62L- CD20+62L+	Replace CD19 (3 new pops)
22	CD45RA/SR181	CD20/SR160	CD20+45RA- CD20+45RA+	Replace CD19 (2 new pops)
Monocyte subsets (also B, CD4 T)				
	Cy5	**Cy5.5**	**Potential populations detected by this combination of stains**	**Comments**
23	HLA PAN/SR151	CD20/SR160	Total PAN II+ Total CD20+ PAN II+20+ (B) PAN II+20- (mono)	New PAN II (3 new pops)
24	HLA 2DR/SR147	CD20/SR160	Total DR+ Total CD20+ DR+20+ (B) DR+20- (mono)	New DR (3 new pops)
25	HLA PAN/SR151	CD4/SR051	Total PAN II+ *Total CD4+* PAN II+4+ (mono) PAN II+4- (B) PAN II-4+(T)	New PAN II (3 new pops)

TABLE 3 (CONTINUED)

26	HLA 2 DR/SR14 7	CD4/SR05 1	Total DR+ Total CD4+ DR+4+ (mono) DR+4- (B) DR-4+ (T)	New DR (3 new pops)
27	CD33/SR 094	CD4/SR05 1	Total CD4+ Total CD33+ CD33+4+ (mono) CD33-4+ (CD4 T)	New way to detect mono (3 new pops)
28	CD14/SR 179	CD4/SR05 1	Total CD4+ Total CD14+ CD4+14+ (mono)	New CD14 (2 new pops)
29	CD14/SR 179	CD3/SR05 5	Total CD14+ Total CD3+	Confirmation of CD3 and CD14 subsets (no new)
30	CD80/SR 101	CD33/SR1 06	Total CD33+ CD33+80- CD33+80+	(2 new pops)
31	CD86/SR 143	CD33/SR1 06	Total CD33+ CD33+86- CD33+86+	(2 new pops)
32	CD45RA/ SR181	CD33/SR1 06	Total CD33+ CD33+45RA- CD33+45RA+	(2 new pops)
33	CD62L/S R098	CD33/SR1 06	Total CD33+ CD33+62L- CD33+62L+	(2 new pops)
		Granulocyte subsets		
Combo #	**Cy5**	**Cy5.5**	**Potential populations detected by this combination of stains**	**Comments**
34	CD16/SR 065	CD45/SR1 39	Total CD45+ (total wbc) Total CD16+ Large, small CD45+16+ Large, small CD45+16-	In Pro-5001 (6 new pops)
35	CD16/SR 065	CD11b/SR 070	Total CD11b+ Total CD16+ Large, small CD16+11b+ Large, small CD16-11b+	In Pro-5001 (4 new pops)
36	CD62L/S R098	CD45/SR1 39	Total CD45 Large, small CD45+62L+ Large, small CD45+62L-	New gran combo (4 new pops)
37	CD11b/S R102	CD45/SR1 39	Total CD45 Large, small CD45+11b+ Large, small CD45+11b-	New gran combo (4 new pops)
38	CD45RB/ SR144	CD4/SR05 1	Total CD4+ Total CD45RB+ CD4+45RB+ CD4-45RB+	New CD45RB (3 new pops?)

TABLE 3 (CONTINUED)

Others				
39	none	none	TCC	1 new pop – total cells
40	none	none	DCC	1 new pop – dead cells

* This is an example of possible cell populations to monitor. Alternative and/or additional populations could be monitored.

TABLE 4

Pilot study- linear discriminant analysis

Best parameters for distinguishing RA and Blood Bank samples in data set

Samples, n=51, Blood bank = 26, RA = 25

Incorrect sample assignment	Single Marker
Single Markers	
7	TCR-αβ T cells as a % of leukocytes
9	CD7 cells as a % of leukocytes
	CD3 cells as a % of leukocytes
10	CD5 cells as a % of leukocytes
	CD4 T cells as a % of leukocytes
11	CD8 T cells as a % of leukocytes
12	CD27 T cells as a % of leukocytes
13	CD16 cells as a % of leukocytes
	CD45 cells (all leukocytes)
	CD8 T cells as a % of leukocytes
14	CD20 intensity on B cells
Marker Pairs	
5	CD4 T cells as a % of leukocytes CD8 T cells as a % of leukocytes
	CD20 intensity on B cells CD7 cells as a % of leukocytes
6	CD45 cells (all leukocytes) CD8 T cells as a % of leukocytes
	CD20 T cells as a % of leukocytes CD7 T cells as a % of leukocytes

Most measurements are averages from 2 to 6 assays

TABLE 5

Information on reagent combinations in panel for Pro-5003

No.	Cy5	Cy5.5	Populations	Comments
001	CD2/SR306	CD4/SR051	3 populations CD2+4+ CD2-4+ CD2+4-	
002	CD2/SR306	CD8/SR212	4 populations CD2+8+bright CD2+8+dull CD2+8+total CD2+8-	
003	CD3/SR054	CD4/SR051	3 populations CD3+4+ CD3+4- CD3-4+	Pro-5002
004	CD3/SR054	CD8/SR212	3 populations CD3+8+ CD3+8- CD3-8+	Pro-5002
005	CD7/SR208	CD4/SR051	3 populations CD7+4+ CD7+4- CD7-4+	Pro-5002
006	CD7/SR208	CD8/SR212	3 populations CD7+8+ CD7+8- CD7-8+	Pro-5002
007	αβTCR/SR0 89	CD7/SR211	5 populations αβ+7+ αβ-7+bright αβ-7+dull αβ-7+total αβ+7-	Pro-5002
008	CD27/SR162	CD4/SR051	3 populations CD27+4+ CD27+4- CD27-4+	Pro-5002
009	CD27/SR162	CD8/SR212	7 populations CD27+8+bright CD27+8+dull CD27+8+total CD27+8- CD27-8+bright CD27-8+dull CD27-8+total	Pro-5002

TABLE 5 (CONTINUED)

010	CD6/SR364	CD4/SR051	3 populations CD6+4- CD6+4+ CD6-4+	
011	CD6/SR364	CD8/SR212	7 populations CD6+8+bright CD6+8+dull CD6+8+total CD6+8- CD6-8+bright CD6-8+dull CD6-8+total	
012	CD26/SR363	CD4/SR051	3 populations CD26+4- CD26+4+ CD26-4+	
013	CD26/SR363	CD8/SR212	5 populations CD26+8- CD26+8+ CD26-8+bright CD26-8+dull CD26-8+total	
014	CD57/SR197	CD4/SR051	2 populations CD57+4+ CD57-4+	
015	CD57/SR197	CD8/SR212	6 populations CD57+8+bright CD57+8+dull CD57+8+total CD57-8+bright CD57-8+dull CD57+8+total	
016	NKB1/SR37 5	CD6/SR362	2 populations NKB1+6+ NKB1-6+	
017	CD45RA/SR 346	CD4/SR051	3 populations CD45RA-4+ CD45RA+4+ CD45RA+4-	Pro-5002
018	CD45RA/SR 346	CD8/SR212	3 populations CD45RA-8+ CD45RA+8+ CD45RA+8-	Pro-5002
019	CD62L/SR2 27	CD4/SR051	3 populations CD62L+4+bright CD62L-4+bright CD62L+4+dull	Pro-5002

TABLE 5 (CONTINUED)

020	CD62L/SR227	CD8/SR212	6 populations CD62L+8+bright CD62L+8+dull CD62L+8+total CD62L-8+bright CD62L-8+dull CD62L-8+total	Pro-5002
021	CD69/SR099	CD4/SR051	2 populations CD69+4+ CD69-4+	Use lysed (5x)
022	CD69/SR099	CD8/SR212	2 populations CD69+8+ CD69-8+	Use lysed (10x)
023	CD25/SR231	CD4/SR051	2 populations CD25+4+ CD25-4+	Use lysed (5x)
024	CD25/SR231	CD8/SR212	2 populations CD25+8+ CD25-8+	Use lysed (10x)
025	TCR-VB3/SR215	CD8/SR212	3 populations TCRVB3+8+ TCRVB3-8+ TCRVB3+8-	Use lysed (10x)
026	TCR-VB5/SR216	CD8/SR212	3 populations TCRVB5+8+ TCRVB5-8+ TCRVB5+8-	Use lysed (10x)
027	TCR-VB8/SR217	CD8/SR212	3 populations TCRVB8+8+ TCRVB8-8+ TCRVB8+8-	Use lysed (10x)
028	NKB1/SR375	CD4/SR051	3 populations NKB1+4+ NKB1+4- NKB1-4+	Use lysed (5x)
029	NKB1/SR375	CD8/SR212	3 populations NKB1+8+ NKB1+8- NKB1-8+	Use lysed (10x)
030	CD5/SR297	CD19/SR050	3 populations CD5-19+ CD5+19+ CD5+19-	Pro-5002
031	CD6/SR364	CD19/SR050	3 populations CD6+19- CD6+19+ CD6-19+	

TABLE 5 (CONTINUED)

032	CD27/SR162	CD19/SR050	3 populations CD27-19+ CD27+19+ CD27+19-	
033	CD2/SR306	CD19/SR050	3 populations CD2+19- CD2-19+ CD2+19+	Stanford study
034	CD80/SR228	CD19/SR050	2 populations CD80+19+ CD80-CD19+	Use lysed (10x)
035	CD86/SR236	CD19/SR050	2 populations CD86+19+ CD86-19+	Use lysed (10x)
036	CD25/SR231	CD19/SR050	2 populations CD25+19+ CD25-19+	Use lysed (10x)
037	CD69/SR099	CD19/SR050	2 populations CD69+19+ CD69-19+	Use lysed (10x)
038	CD62L/SR2 27	CD20/SR224	1 population CD62L+20+	Pro-5002
039	CD45RA/SR 346	CD20/SR224	2 populations CD45RA+20+ CD45RA+20-	Pro-5002
040	HLA PAN/SR229	CD20/SR224	2 populations HLAPAN II+20+ HLAPAN II+20-	Pro-5002
041	HLA 2DR/SR230	CD20/SR224	2 populations HLADR+20+ HLADR+20-	Pro-5002
042	HLA DP/SR370	CD20/SR224	2 populations HLADP+20+ HLADP+20-	
043	HLA PAN/SR229	CD4/SR051	3 populations HLAPAN+4+ HLAPAN+4- HLAPAN-4+	Pro-5002
044	HLA 2DR/SR230	CD4/SR051	3 populations HLADR+4+ HLADR+4- HLADR-4+	Pro-5002
045	HLA DP/SR370	CD4/SR051	3 populations HLADP+4+ HLADP+4- HLADP-4+	

TABLE 5 (CONTINUED)

046	CD33/SR232	CD14/SR366	4 populations CD33+14+ CD33+14+total CD33dull14+ CD33+14-	Good crosscheck on CD33 use doped down CD14
047	CD33/SR232	CD4/SR051	3 populations CD33+4+ CD33-4+ CD33+4-	Pro-5002 Second check on CD33
048	CD16/SR065	CD14/SR366	5 populations CD16+14+bright CD16+14+dull CD16+14+total CD16-14+ CD16+14-	
049	CD64/SR182	CD4/SR051	2 populations CD64+4+ CD64-4+	
050	CD64/SR182	CD16/SR072	3 populations CD64+16+ CD64+16- CD64-16+	Stanford study
051	CD45RA/SR 346	CD14/SR366	2 populations CD45RA+14+ CD45RA+14-	Doped down CD14
052	CD62L/SR2 27	CD14/SR366	1 population CD62L+14+	Doped down CD14
053	CD86/SR236	CD14/SR366	1 population CD86+14+	Doped down CD14
054	CD45/SR132	CD14/SR366	4 populations CD45+14-total CD45bright14- CD45dull14- CD45+14+	Nice breakdown of lymphs, grans, mono Stanford study Lysed, 1:4 diluted (0.5x)
055	CD45/SR132	CD16/SR072	3 populations + total CD45+ CD45+16+hi_sl CD45+16+lo_sl CD45+16-	Pro-5002 1:4 diluted blood
056	CD15/SR195	CD16/SR072	2 populations CD15+16+ CD15-16+	1:4 diluted blood
057	CD18/SR374	CD15/SR372	2 populations CD18+15+ CD18+15-	1:4 diluted blood

TABLE 5 (CONTINUED)

058	CD45/SR132	CD14/SR366	4 populations CD45+14-total CD45bright14- CD45dull14- CD45+14+	Nice breakdown of lymphs, grans, mono Stanford study 1:4 diluted blood
059	CD11b/SR063	CD15/SR372	2 populations CD11b+15+ CD11b+15-	1:4 diluted blood
060	CD32/SR180	CD15/SR372	2 populations CD32+15+ CD32+15-	1:4 diluted blood

TABLE 6

Soluble Factor Immunoassays

No	ASSAY
1	IL-1 alpha
2	IL-1 beta
3	IL-1ra
4	IL-1sRI
5	IL-1sRII
6	IL-6
7	IL-8
8	IL-10
9	RF (all isotypes)
10	RF IgM
11	RF IgG
12	RF IgA
13	CRP
14	SAA
15	MMP-3
16	MMP-9
17	TIMP-1
18	TNF alpha
19	INF gamma
20	TGF beta
21	sCD62E
22	sCD62P

No	ASSAY
23	IL-2
24	IL-3
25	IL-4
26	IL-5
27	MMP-1
28	MMP-2
29	MMP13
30	TIMP-2
31	TIMP-3
32	sCD44
33	ScD54 ICAM-1
34	sCD62L
35	RANTES
36-51	Immunoglobulin H and L isotypes (16 assays)

S = soluble

TABLE 7

Assay Numbers	Dye	Antigen	SR###	Dye	Antigen	SR###	Dye	Antigen	SR###	Format
General										
ASY3-149	Cy5	CD45	SR712	Cy5.5	CD14	SR503	Cy7-APC	CD16	SR433	WB 0.25x
ASY3-132	Cy5	CD45	SR712	Cy5.5	CD14	SR503	Cy7-APC	CD16	SR433	Lysed 0.25x
T Cells (all or both 4 and 8)										
ASY3-001	Cy5	CD4	SR349	Cy5.5	CD8	SR212	Cy7-APC	CD3	SR435	WB
ASY3-150	Cy5	CD2	SR306	Cy5.5	CD4	SR506	Cy7-APC	CD8	SR529	WB
ASY3-066	Cy5	TCRαβ	SR660	Cy5.5	TCRγδ	SR663	Cy7-APC	CD3	SR435	WB
ASY3-151	Cy5	TCRαβ	SR660	Cy5.5	CD4	SR506	Cy7-APC	CD8	SR529	WB
CD8 Cells										
ASY3-055	Cy5	CD62L	SR227	Cy5.5	CD45RA	SR453	Cy7-APC	CD8	SR529	WB
ASY3-008	Cy5	CD57	SR342	Cy5.5	CD6	SR362	Cy7-APC	CD8	SR529	WB
ASY3-152	Cy5	CD27	SR225	Cy5.5	CD45RA	SR453	Cy7-APC	CD8	SR529	WB
ASY3-178	Cy5	CD28	SR675	Cy5.5	CD62L	SR454	Cy7-APC	CD8	SR529	WB
ASY3-179	Cy5	CD28	SR675	Cy5.5	CD45RA	SR453	Cy7-APC	CD8	SR529	WB
ASY3-079	Cy5	CD69	SR235	Cy5.5	CD25	SR616	Cy7-APC	CD8	SR529	Lysed 5x
ASY3-080	Cy5	CD71	SR654	Cy5.5	CD57	SR619	Cy7-APC	CD8	SR529	Lysed 5x
ASY3-089	Cy5	CD38	SR671	Cy5.5	CD72	SR592	Cy7-APC	CD8	SR529	WB
ASY3-090	Cy5	CD28	SR675	Cy5.5	CD26	SR343	Cy7-APC	CD8	SR529	WB
ASY3-091	Cy5	CCR5	SR502	Cy5.5	CD8	SR212	Cy7-APC	CD3	SR435	WB
ASY3-142	Cy5	CD4	SR349	Cy5.5	CD7	SR211	Cy7-APC	CD8	SR529	WB
ASY3-145	Cy5	CD44	SR558	Cy5.5	CD7	SR211	Cy7-APC	CD8	SR529	WB-0.25x

TABLE 7 (CONTINUED)

Assay Numbers	Dye	Antigen	SR###	Dye	Antigen	SR###	Dye	Antigen	SR###	Format
CD4 T Cells										
ASY3-056	Cy5	CD62L	SR227	Cy5.5	CD45RA	SR453	Cy7-APC	CD4	SR530	WB
ASY3-004	Cy5	CD4	SR349	Cy5.5	CD27	SR161	Cy7-APC	CD3	SR435	WB
ASY3-153	Cy5	CD26	SR363	Cy5.5	CD4	SR506	Cy7-APC	CD3	SR435	WB
ASY3-154	Cy5	CD57	SR342	Cy5.5	CD4	SR506	Cy7-APC	CD3	SR435	WB
ASY3-155	Cy5	CD62L	SR227	Cy5.5	CD4	SR506	Cy7-APC	CD3	SR435	WB
ASY3-188	Cy5	CD27	SR225	Cy5.5	CD45RA	SR453	Cy7-APC	CD4	SR530	WB
ASY3-180	Cy5	CD28	SR675	Cy5.5	CD45RA	SR453	Cy7-APC	CD4	SR530	WB
ASY3-063	Cy5	CD7	SR208	Cy5.5	CD6	SR362	Cy7-APC	CD4	SR530	WB
ASY3-156	Cy5	CD44	SR558	Cy5.5	CD4	SR506	Cy7-APC	CD3	SR435	WB-0.25x
ASY3-157	Cy5	CD89	SR447	Cy5.5	CD4	SR506	Cy7-APC	CD3	SR435	WB-0.25x
CD4 T and Mono										
ASY3-137	Cy5	CD69	SR235	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	Lysed-2x
ASY3-138	Cy5	CD25	SR231	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	Lysed-2x
ASY3-158	Cy5	CCR5	SR502	Cy5.5	CD4	SR506	Cy7-APC	CD14	SR719	WB
ASY3-159	Cy5	CD38	SR671	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	WB
ASY3-160	Cy5	CD86	SR236	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	WB
ASY3-139	Cy5	CD71	SR654	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	Lysed-2x

TABLE 7 (CONTINUED)

Assay Numbers	Dye	Antigen	SR###	Dye	Antigen	SR###	Dye	Antigen	SR###	Format
B Cells										
ASY3-143	Cy5	CD5	SR297	Cy5.5	CD19	SR050	Cy7-APC	CD20	SR718	WB
ASY3-161	Cy5	CD72	SR100	Cy5.5	CD19	SR050	Cy7-APC	CD20	SR718	WB
ASY3-162	Cy5	CD80	SR228	Cy5.5	CD86	SR706	Cy7-APC	CD20	SR718	WB
ASY3-163	Cy5	CD69	SR235	Cy5.5	CD71	SR655	Cy7-APC	CD20	SR729	Lysed-5x
B Cell and Mono										
ASY3-164	Cy5	HLADP	SR370	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR718	WB
ASY3-165	Cy5	HLADQ	SR500	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR718	WB
ASY3-166	Cy5	HLADR	SR230	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR718	WB
ASY3-167	Cy5	HLAPAN	SR229	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR718	WB
ASY3-168	Cy5	CD14	SR179	Cy5.5	CD45RA	SR453	Cy7-APC	CD20	SR729	WB
ASY3-169	Cy5	CD40	SR634	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR718	WB
ASY3-170	Cy5	CD62L	SR227	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR729	WB
Monocyte										
ASY3-171	Cy5	CD33	SR232	Cy5.5	CD14	SR503	Cy7-APC	CD4	SR530	WB
ASY3-172	Cy5	CD4	SR349	Cy5.5	CD11b	SR371	Cy7-APC	CD14	SR719	WB-0.25x
ASY3-045	Cy5	CD16b	SR359	Cy5.5	CD66b	SR536	Cy7-APC	CD16	SR433	WB-0.25x
ASY3-173	Cy5	CD64	SR365	Cy5.5	CD14	SR503	Cy7-APC	CD16	SR433	WB-0.25x
ASY3-049	Cy5	CD32	SR379	Cy5.5	CD15	SR372	Cy7-APC	CD16	SR433	WB-0.25x
ASY3-047	Cy5	CD18	SR374	Cy5.5	CD11b	SR371	Cy7-APC	CD16	SR433	WB-0.25x
ASY3-174	Cy5	CD44	SR558	Cy5.5	CD15	SR372	Cy7-APC	CD14	SR719	WB-0.25x
ASY3-175	Cy5	CD89	SR658	Cy5.5	CD15	SR372	Cy7-APC	CD14	SR719	WB-0.25x

TABLE 7 (CONTINUED)

Assay Numbers	Dye	Antigen	SR###	Dye	Antigen	SR###	Dye	Antigen	SR###	Format
ASY3-128	Cy5	CD9	SR310	Cy5.5	CD15	SR372	Cy7-APC	CD16	SR433	Lysed 0.25x
ASY3-148	Cy5	CD123	SR289	Cy5.5	CD32	SR704	Cy7-APC	CD16	SR433	WB 0.25x
ASY3-147	Cy5	CD123	SR289	Cy5.5	CD15	SR372	Cy7-APC	CD16	SR433	WB 0.25x
NK										
ASY3-038	Cy5	NKB1	SR375	Cy5.5	CD5	SR298	Cy7-APC	CD7	SR490	WB
ASY3-071	Cy5	CD57	SR342	Cy5.5	CD5	SR298	Cy7-APC	CD7	SR490	WB
ASY3-085	Cy5	CD56	SR676	Cy5.5	CD2	SR352	Cy7-APC	CD3	SR435	WB
ASY3-086	Cy5	CD56	SR676	Cy5.5	CD5	SR298	Cy7-APC	CD7	SR490	WB
Controls										
ASY3-050	Cy5	MOPC	SR344	Cy5.5	MOPC	SR350	Cy7-APC	MOPC	SR624	WB
ASY3-176	Cy5	CD5	SR297	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR729	WB
ASY3-177	Cy5	CD5	SR297	Cy5.5	CD14	SR503	Cy7-APC	CD20	SR729	Lysed -1x
ASY3-082	Cy5	CD4	SR349	Cy5.5	CD8	SR212	Cy7-APC	CD3	SR435	Lysed - 1x

TABLE 8
BLOOD SAMPLING/DOSING SCHEDULE

DAY	DAY	DAY	DAY	DAY	DAY	DAY
Fri	Mon	Tue	Wed	Thurs	Fri	Thurs
-3	0	1	2	3	4	10
Screening						
-	Aspirin	Aspirin	Aspirin	Aspirin	-	-
Blood	Blood		Blood	Blood	Blood	Blood

Blood draw between 8 and 9 am each day

TABLE 9

		Ch0			Ch1			Ch2			
	Assay #	Dye	Antigen	SR###	Dye	Antigen	SR###	Dye	Antigen	SR###	Format
1	ASY3-102	Cy5	CD36	SR679	Cy5.5	CD9	SR678	Cy7 APC	CD61	SR641	WB
2	ASY3-103	Cy5	CD42a	SR684	Cy5.5	CD41a	SR683	Cy7 APC	CD61	SR641	WB
3	ASY3-104	Cy5	CD42a	SR684	Cy5.5	CD62p	SR590	Cy7 APC	CD61	SR641	WB
4	ASY3-105	Cy5	CD42b	SR685	Cy5.5	CD41a	SR683	Cy7 APC	CD61	SR641	WB
5	ASY3-106	Cy5	CD42b	SR685	Cy5.5	CD62p	SR590	Cy7 APC	CD61	SR641	WB
6	ASY3-107	Cy5	CD62p	SR686	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	WB
7	ASY3-108	Cy5	CD63	SR687	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	WB
8	ASY3-109	Cy5	PAC-1	SR673	Cy5.5	CD9	SR678	Cy7 APC	CD61	SR641	WB
9	ASY3-110	Cy5	CD29	SR150	Cy5.5	CD9	SR678	Cy7 APC	CD41a	SR640	WB
10	ASY3-111	Cy5	CD62p	SR686	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	TRT
11	ASY3-112	Cy5	CD63	SR687	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	TRT
12	ASY3-113	Cy5	PAC-1	SR673	Cy5.5	CD9	SR678	Cy7 APC	CD61	SR641	TRT
13	ASY3-114	Cy5	CD42b	SR685	Cy5.5	CD62p	SR590	Cy7 APC	CD61	SR641	TRT
14	ASY3-115	Cy5	CD62p	SR686	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	NTRT

TABLE 9 (CONTINUED)

15	ASY3-116	Cy5	CD63	SR687	Cy5.5	CD61	SR681	Cy7 APC	CD41a	SR640	NTRT
16	ASY3-117	Cy5	PAC-1	SR673	Cy5.5	CD9	SR678	Cy7 APC	CD61	SR641	NTRT
17	ASY3-118	Cy5	CD42b	SR685	Cy5.5	CD62p	SR590	Cy7 APC	CD61	SR641	NTRT
18	ASY3-066	Cy5	TCRab	SR660	Cy5.5	TCRgd	SR663	Cy7APC	CD3	SR435	WB
19	ASY3-151	Cy5	TCRab	SR660	Cy5.5	CD4	SR506	Cy7APC	CD8	SR529	WB
20	ASY3-055	Cy5	CD62L	SR227	Cy5.5	CD45RA	SR453	Cy7APC	CD8	SR529	WB
21	ASY3-178	Cy5	CD28	SR675	Cy5.5	CD62L	SR454	Cy7APC	CD8	SR529	WB
22	ASY3-091	Cy5	CCR5	SR502	Cy5.5	CD8	SR212	Cy7APC	CD3	SR435	WB
23	ASY3-142	Cy5	CD4	SR349	Cy5.5	CD72	SR211	Cy7APC	CD8	SR529	WB
24	ASY3-056	Cy5	CD62L	SR227	Cy5.5	CD45RA	SR453	Cy7APC	CD4	SR530	WB
25	ASY3-180	Cy5	CD28	SR675	Cy5.5	CD45RA	SR453	Cy7APC	CD4	SR530	WB
26	ASY3-158	Cy5	CCR5	SR502	Cy5.5	CD4	SR506	Cy7APC	CD14	SR719	WB
27	ASY3-160	Cy5	CD86	SR236	Cy5.5	CD14	SR503	Cy7APC	CD4	SR530	WB
28	ASY3-186	Cy5	CD5	SR297	Cy5.5	CD19	SR050	Cy7APC	CD20	SR729	WB
29	ASY3-181	Cy5	CD80	SR228	Cy5.5	CD86	SR706	Cy7-APC	CD20	SR729	WB
30	ASY3-182	Cy5	HLADP	SR370	Cy5.5	CD14	SR503	Cy7APC	CD20	SR729	WB
31	ASY3-183	Cy5	HLADQ	SR500	Cy5.5	CD14	SR503	Cy7APC	CD20	SR729	WB
32	ASY3-184	Cy5	HLADR	SR230	Cy5.5	CD14	SR503	Cy7APC	CD20	SR729	WB
33	ASY3-185	Cy5	CD40	SR634	Cy5.5	CD14	SR503	Cy7APC	CD20	SR729	WB
34	ASY3-171	Cy5	CD33	SR232	Cy5.5	CD14	SR503	Cy7APC	CD4	SR530	WB
35	ASY3-038	Cy5	NKB1	SR375	Cy5.5	CD5	SR298	Cy7APC	CD7	SR490	WB

TABLE 9 (CONTINUED)

36	ASY3-085	Cy5	CD56	SR676	Cy5.5	CD2	SR352	Cy7APC	CD3	SR435	WB
37	ASY3-050	Cy5	MOPC	SR344	Cy5.5	MOPC	SR350	Cy7APC	MOPC	SR624	WB
38	ASY3-156	Cy5	CD44	SR558	Cy5.5	CD4	SR506	Cy7APC	CD3	SR435	WB-0.25x
39	ASY3-045	Cy5	CD16b	SR359	Cy5.5	CD66b	SR536	Cy7APC	CD16	SR433	WB-0.25x
40	ASY3-173	Cy5	CD64	SR365	Cy5.5	CD14	SR503	Cy7APC	CD16	SR433	WB-0.25x
41	ASY3-148	Cy5	CD123	SR289	Cy5.5	CD32	SR704	Cy7-APC	CD16	SR433	WB 0.25x
42	ASY3-147	Cy5	CD123	SR289	Cy5.5	CD15	SR372	Cy7-APC	CD16	SR433	WB 0.25x

Claims

1. A biological marker identification system comprising:

a) an integrated database comprising a plurality of data categories, said data categories comprising,

i) levels of a plurality of cell populations and/or cell associated molecules in the biological fluid of an organism, and/or levels of a plurality of soluble factors in the biological fluid of an organism, and

ii) information associated with a plurality of clinical parameters of an organism;

b) data from a plurality of organisms corresponding to said data categories; and

i) processing means for correlating data within the data categories, wherein correlation analysis of data categories can be made to identify the data category or categories indicating normal biologic processes, pathogenic processes, or pharmacological responses to therapeutic intervention,

wherein said identified category or categories are biological markers.

2. The biological marker identification system of claim 1 wherein said data for levels of cell populations and/or cell associated molecules are obtained by microvolume laser scanning cytometry.

3. The biological marker identification system of claims 1 and 2 comprising at least 20 cell population and/or cell associated molecules level data categories.

4. The biological marker identification system of claim 3 comprising at least 30 cell population and/or cell associated molecules level data categories.

5. The biological marker identification system of claim 3 comprising at least 40 cell population and/or cell associated molecules level data categories.

6. The biological marker identification system of claims 1-3 wherein the soluble factor is a soluble protein.

7. The biological marker identification system of claim 1 wherein the soluble factor is a small molecule.

8. The biological marker identification system of claim 1 wherein said data for levels of soluble factors are obtained by microvolume laser scanning cytometry.

9. The biological marker identification system of claim 1 wherein said data for levels of soluble factors are obtained by immunoassays.

10. The biological marker identification system of claim 1 comprising at least 20 soluble factor level data categories.

11. The biological marker identification system of claim 10 comprising at least 30 soluble factor level data categories.

12. The biological marker identification system of claim 10 comprising at least 40 soluble factor level data categories.

13. The biological marker identification system of claim 1 wherein data from at least some of said organisms is included a plurality of times.

14. The biological marker identification system of claim 1 wherein said data categories further include:

iii) genotype information associated with an organism.

15. The biological marker identification system of claim 1 wherein said data for levels of soluble factors are obtained by mass spectrometry.

16. The biological marker identification system of claim 1 wherein said information associated with said clinical parameters is selected from the group consisting of age, gender, weight, height, body type, medical history, family history, environmental factors and manifestation and categorization of disease or medical condition.

17. The biological marker identification system of claim 1 wherein said data is obtained from organisms prior to and after the administration of a therapeutic treatment.

18. The biological marker identification system of claim 1 wherein at least some of said data is obtained from an organism having been previously diagnosed as having a predetermined disease or medical condition.

19. The biological marker identification system of claim 1 wherein at least some of said data is obtained at a plurality of times from an organism having been previously diagnosed as having a predetermined disease or medical condition.

20. The biological marker identification system of claims 18 and 19 wherein said predetermined disease or medical condition is rheumatoid arthritis.

21. The biological marker identification system of claims 18 and 19 wherein said predetermined disease or medical condition is selected from the group consisting of rheumatoid arthritis, asthma, allergy and multiple sclerosis.

22. The biological marker identification system of claim 1 wherein said data categories comprise levels of a plurality of cell populations and/or cell associated molecules in the biological fluid of an organism and levels of a plurality of soluble factors in the biological fluid of an organism.

23. A method for identifying a biological marker for a given disease or medical condition comprising:

correlating information obtained from a plurality of organisms, at least some of said organisms having said disease or medical condition, wherein information is associated with a plurality of data categories, and wherein said data categories comprise,

i) levels of a plurality of cell populations and/or cell associated molecules in the biological fluid of an organism, and/or levels of a plurality of soluble factors in the biological fluid of an organism, and

ii) information associated with a plurality of clinical parameters of an organism;

identifying a data category where organisms having said disease or medical condition may be differentiated from those organisms not having said disease or medical condition, wherein said identified category is a biological marker for said disease.

24. The method for identifying a biological marker of claim 23 wherein said data for levels of cell populations and/or cell associated molecules are obtained by microvolume laser scanning cytometry.

25. The method for identifying a biological marker of claim 23 comprising at least 20 cell population and/or cell associated molecules level data categories.

26. The method for identifying a biological marker of claim 25 comprising at least 30 cell population and/or cell associated molecules level data categories.

27. The method for identifying a biological marker of claim 25 comprising at least 40 cell population and/or cell associated molecules level data categories.

28. The method for identifying a biological marker of claim 23 wherein said data for levels of soluble factors are obtained by microvolume laser scanning cytometry.

29. The method for identifying a biological marker of claim 23 comprising at least 20 soluble factor level data categories.

30. The method for identifying a biological marker of claim 29 comprising at least 30 soluble factor level data categories.

31. The method for identifying a biological marker of claim 29 comprising at least 40 soluble factor level data categories.

32. The method for identifying a biological marker of claim 23 wherein said data categories further include:

iii) genotype information associated with any organism.

33. The method for identifying a biological marker of claim 23 wherein said data for levels of soluble factors are obtained by mass spectrometry.

34. The method for identifying a biological maker of claim 23 wherein said data for levels of soluble factors are obtained by immunoassays.

35. The method for identifying a biological marker of claim 23 wherein said information associated with said clinical parameters is selected from the group consisting of age, gender, weight, height, body type, medical history, family history, environmental factors and manifestation and categorization of disease or medical condition.

36. The method for identifying a biological marker of claim 23 wherein said disease is rheumatoid arthritis.

37. The method for identifying a biological marker of claim 23 wherein said disease is selected from the group consisting of rheumatoid arthritis, asthma, allergy and multiple sclerosis.

38. A phenotype of an organism comprising a plurality of biological parameters comprising:

i) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

ii) the results of at least 20 assays relating to soluble factors; and

iii) clinical parameters.

39. The phenotype of claim 38 comprising the results of at least 40 assays relating to cell populations and/or cell associated molecules.

40. The phenotype of claim 38 comprising the results of at least 40 assays relating to soluble factors.

41. The phenotype of claim 38 further comprising genotype information of said organism.

42. A phenotype of a class or subclass of organisms comprising a plurality of biological parameters from a plurality of members of said class or subclass; wherein from each said member said biological parameters comprise:

i) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

ii) the results of at least 20 assays relating to soluble factors; and

iii) clinical parameters.

43. The phenotype of claim 42 comprising the results of at least 40 assays relating to cell populations and/or cell associated molecules.

44. The phenotype of claim 42 comprising the results of at least 40 assays relating to soluble factors.

45. The phenotype of claim 42 further comprising genotype information of each said member.

46. A system for creating the phenotype of an organism comprising:

i) obtaining biological parameters from said organism comprising:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters; and

ii) entering said biological parameters into an integrated data base.

47. The system of claim 46 comprising the results of at least 40 assays relating to cell populations and/or cell associated molecules.

48. The system of claim 46 comprising the results of at least 40 assays relating to soluble factors.

49. The system of claim 46 wherein said biological parameters further comprise genotype information.

50. A method for evaluating the effect of a perturbation on an organism comprising:

i) obtaining the phenotype of said organism prior to and after said perturbation; and

ii) comparing the information in said prior to and after phenotypes to identify changed parameters;

wherein said phenotypes are comprised of:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters.

51. The method of claim 50, wherein said phenotypes comprise at least 40 assays relating to cell populations and/or cell associated molecules.

52. The method of claim 50, wherein said phenotypes comprise at least 40 assays relating to soluble factors.

53. The method of claim 50, wherein said phenotypes further comprise genotype information of said organism.

54. A method for evaluating the effect of a perturbation on a class or subclass of organisms comprising:

i) obtaining the phenotype of a plurality of members of said class or subclass of organisms prior to and after said perturbation;

ii) comparing the information in said prior to and after phenotype to identify changed parameters;

wherein said phenotypes are comprised of:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters.

55. A method for evaluating the effect of a perturbation on an organism or class or subclass or organisms comprising:

i) obtaining the phenotype of a plurality of said organisms who have not been effected by said perturbation and the phenotype of one or more of said organisms who have been effected by said perturbation; and

ii) comparing the information in the phenotypes of said plurality of organisms who have not been effected by said perturbation with the phenotype of the one or more organisms who have been effected by said perturbation to identify changed parameters;

wherein said phenotypes are comprised of:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters.

56. A system for the identification of biological markers of a disease or medical condition in an animal model of said disease or medical condition comprising:

a) an integrated database comprising a plurality of data categories, said data categories comprising,

i) levels of a plurality of cell populations and/or cell associated molecules in the biological fluid of an animal, and/or levels of a plurality of soluble factors in the biological fluid of an animal, and

ii) information associated with a plurality of physical parameters of an animal;

b) data from a plurality of animals corresponding to said data categories; and

i) processing means for correlating data within the data categories, wherein correlation analysis of data categories can be made to identify the data category or categories indicating normal biologic processes, pathogenic processes, or pharmacological responses to candidate therapeutic intervention;

wherein said identified category or categories are biological markers.

57. The biological marker identification system of claim 56 wherein said data for levels of cell populations and/or cell associated molecules are obtained by microvolume laser scanning cytometry.

58. The biological marker identification system of claims 56 and 57 comprising at least 20 cell population and/or cell associated molecules level data categories.

59. The biological marker identification system of claim 58 comprising at least 40 cell population and/or cell associated molecules level data categories.

60. The biological marker identification system of claim 56 comprising at least 20 soluble factor level data categories.

61. The biological marker identification system of claim 56 comprising at least 40 soluble factor level data categories.

62. The biological marker identification system of claim 56 wherein said data categories further include:

genotype information associated with an animal.

63. A method for identifying a biological marker for a given disease or medical condition in an animal model of said disease or medical condition comprising:

providing an animal model of said disease or medical condition;

correlating information obtained from a plurality of individual animals, at least some of said individual animals having said disease or medical condition, wherein information is associated with a plurality of data categories, and wherein said data categories comprise,

i) levels of a plurality of cell populations and/or cell associated molecules in the biological fluid of an individual, and/or levels of a plurality of soluble factors in the biological fluid of an individual animal; and

ii) information associated with a plurality of physical parameters of an individual animal;

identifying a data category where individual animals having said disease or medical condition may be differentiated from those individual animals not having said disease or medical condition, wherein said identified category is a biological marker for said disease in said animal model.

64. A method for identifying a biological marker for a given disease or medical in humans, comprising:

providing an animal model of said disease or medical condition;

identifying a biological marker for said disease or medical condition in the animal model of said disease or medical condition according to the method of claim 63; and

determining if said biological marker is diagnostic or prognostic of said disease or medical condition in humans.

65. A method for assaying a candidate therapeutic agent directed against a human disease or medical condition, the method comprising:

providing an animal model of said disease or medical condition;

identifying at least one biological marker of said disease or medical condition in said animal model by the method of claim 63;

treating said animal model with said candidate therapeutic; and

monitoring the response of said biological markers in said animal model.

66. A method for monitoring the results of a clinical study in humans with a given medical disease or condition comprising:

evaluating biological markers in humans that are homologues of biological markers identified in animal models of said medical disease or condition.

67. A method for designing an improved animal model for a human disease or medical condition comprising:

identifying human biological markers relative to said disease or medical condition;

tailoring the animal model to more accurately simulate said disease or medical condition by elevating or reducing the levels of the animal homologues of said human biological marker.

68. A method for identifying an animal model of a disease or medical condition comprising:

i) obtaining the phenotype of a plurality of potential animal models of said disease or medical condition;

ii) obtaining the phenotype of organism having said disease or medical condition;

iii) comparing the potential animal model phenotypes with the phenotype of the organisms having said disease or medical condition to identify the animal model phenotype that most closely simulates the phenotype of the organisms having said disease or medical condition;

wherein said phenotypes are comprised of:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters.

69. The phenotype of claim 38 wherein said organism is selected from the group consisting of a human, an animal, a plant, and a virus.

70. The phenotype of claim 42 wherein said class or subclass or organisms is selected from the group consisting of humans, animals, plants and virus.

71. A method for evaluating the effects of a genetic alteration on a plant or animal comprising:

i) obtaining the phenotype of said plant or animal that has been genetically altered and the phenotype of the non-genetically altered plant or animal; and

ii) comparing the information in the genetically-altered and non-genetically altered phenotype to identify changed parameters;

wherein said phenotypes are comprised of:

a) the results of at least 20 assays relating to cell populations and/or cell associated molecules;

b) the results of at least 20 assays relating to soluble factors; and

c) clinical parameters.

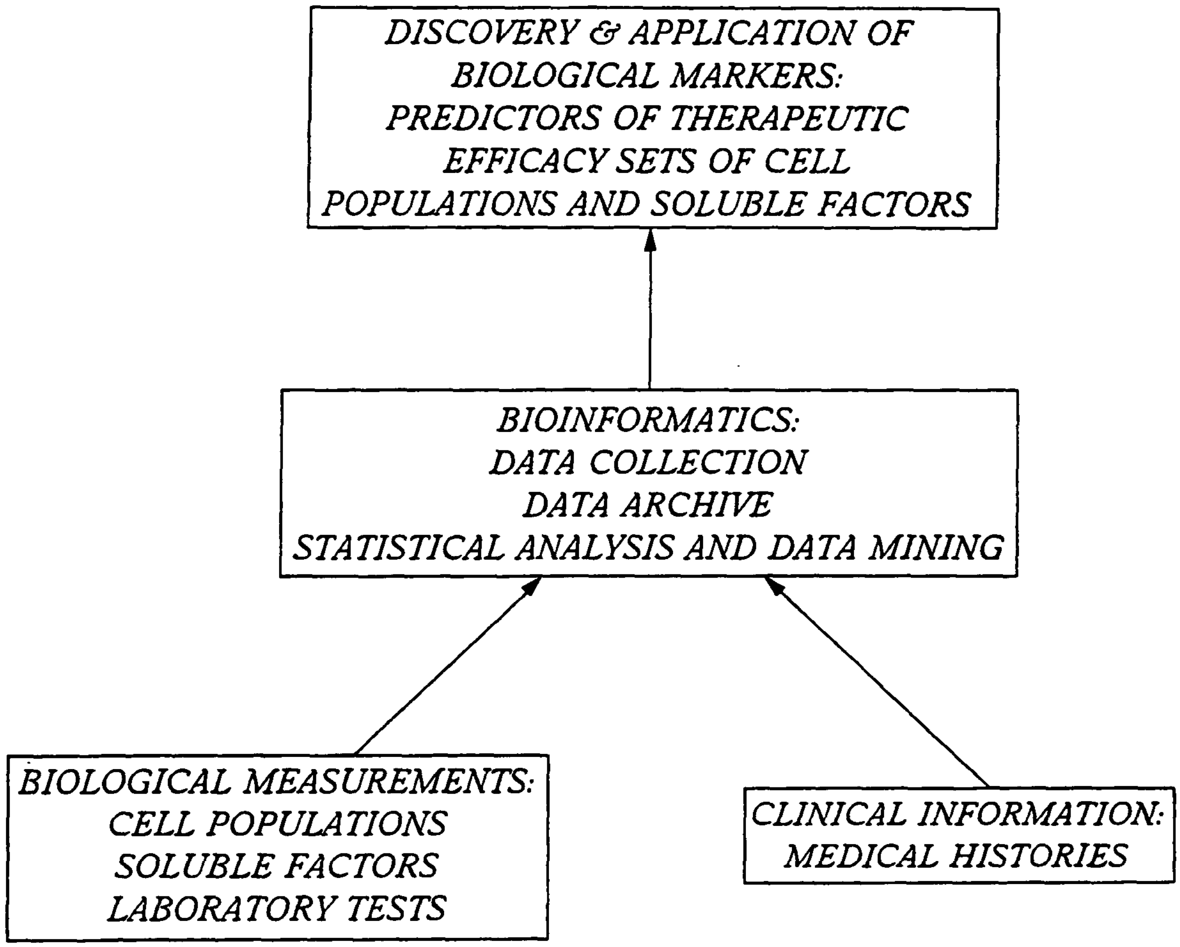

DISCOVERY & APPLICATION OF BIOLOGICAL MARKERS: PREDICTORS OF THERAPEUTIC EFFICACY SETS OF CELL POPULATIONS AND SOLUBLE FACTORS
BIOINFORMATICS: DATA COLLECTION DATA ARCHIVE STATISTICAL ANALYSIS AND DATA MINING
BIOLOGICAL MEASUREMENTS: CELL POPULATIONS SOLUBLE FACTORS LABORATORY TESTS
CLINICAL INFORMATION: MEDICAL HISTORIES

Figure 1

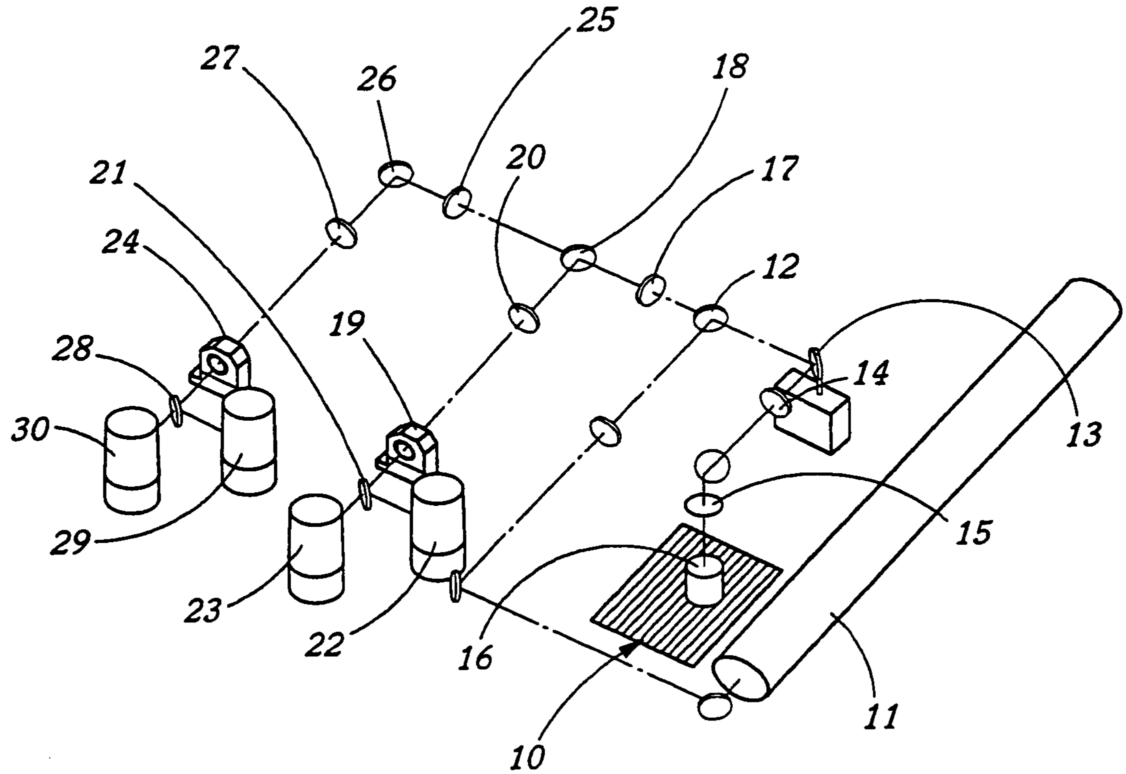


Figure 2

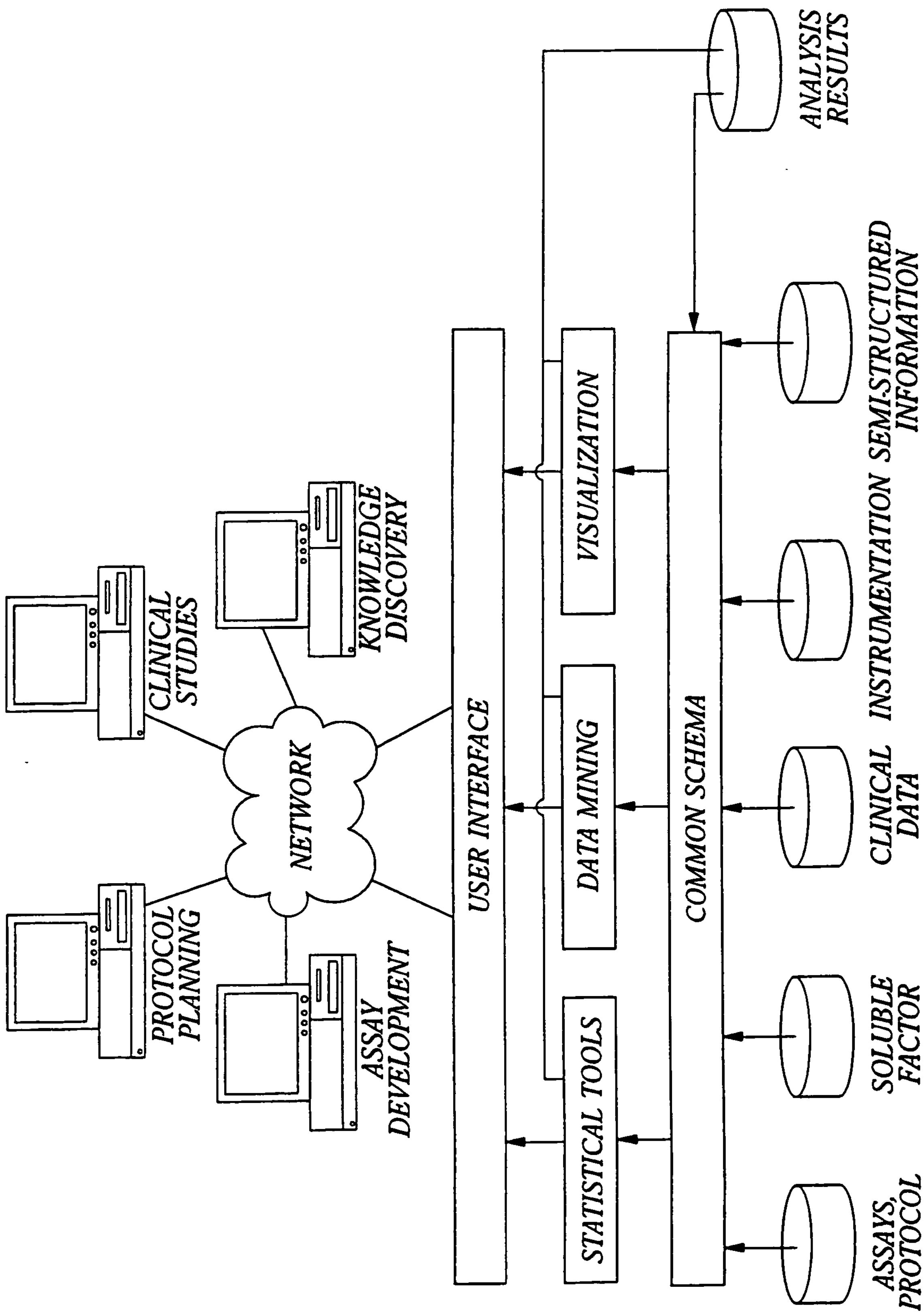

ANALYSIS RESULTS
USER INTERFACE
VISUALIZATION
DATA MINING
STATISTICAL TOOLS
COMMON SCHEMA
SEMISTRUCTURED INFORMATION
INSTRUMENTATION
CLINICAL DATA
SOLUBLE FACTOR
ASSAYS, PROTOCOL
NETWORK
CLINICAL STUDIES
KNOWLEDGE DISCOVERY
PROTOCOL PLANNING
ASSAY DEVELOPMENT

Figure 3

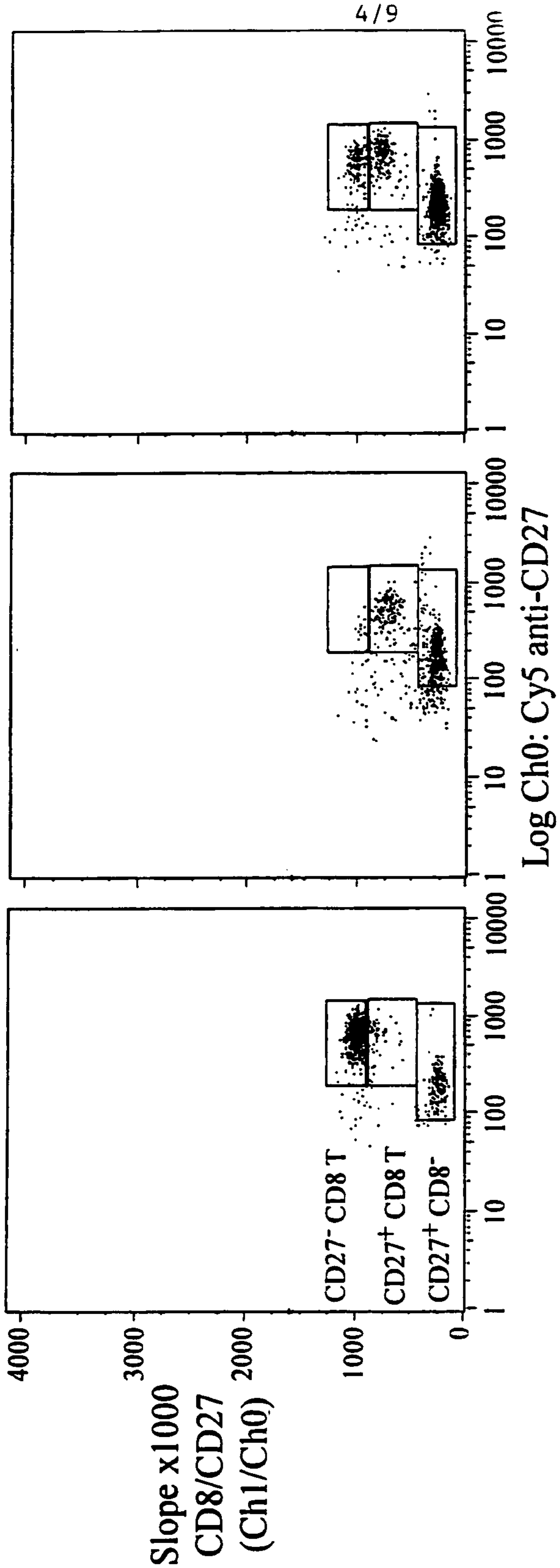


Figure 4A

Figure 4B

Figure 4C

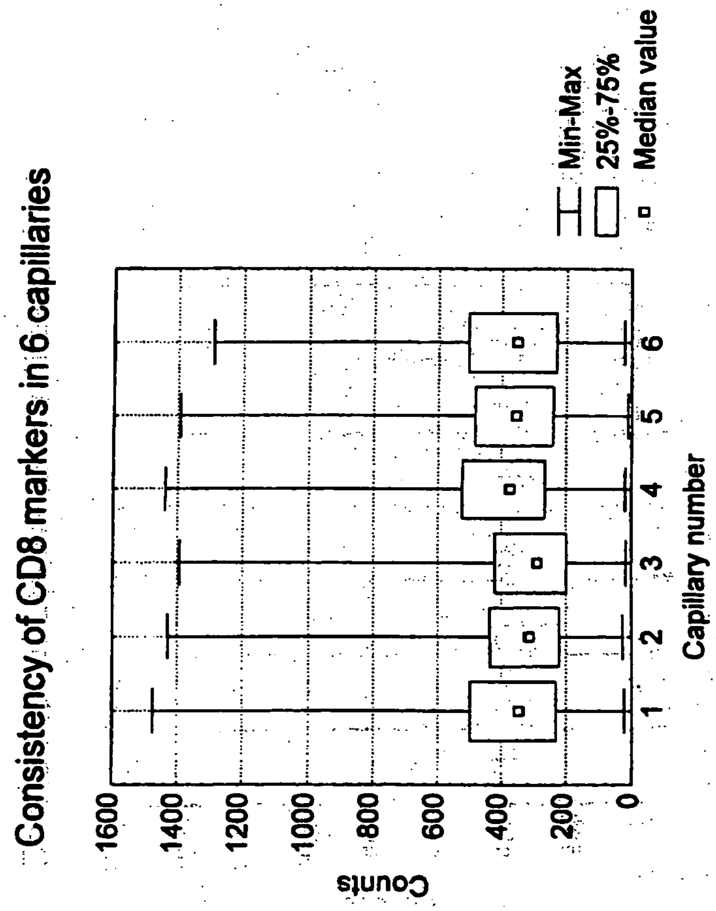


Figure 5A

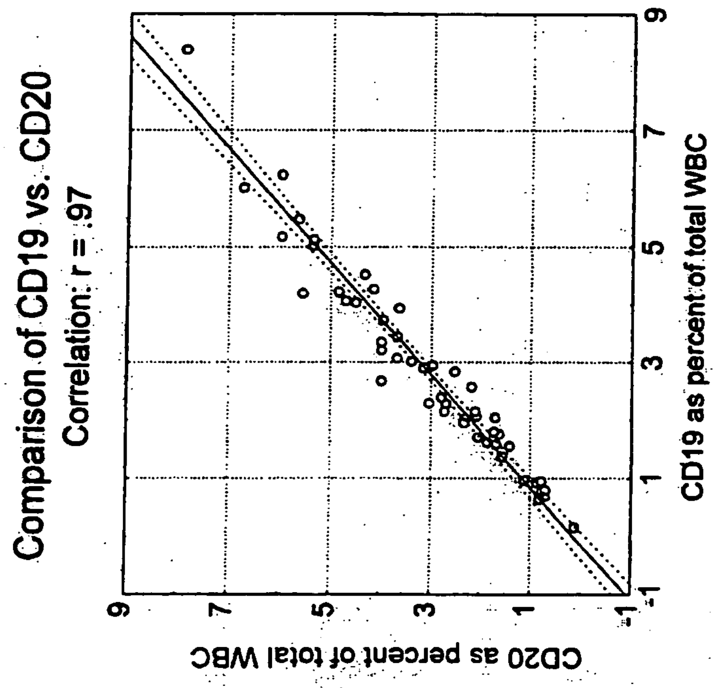
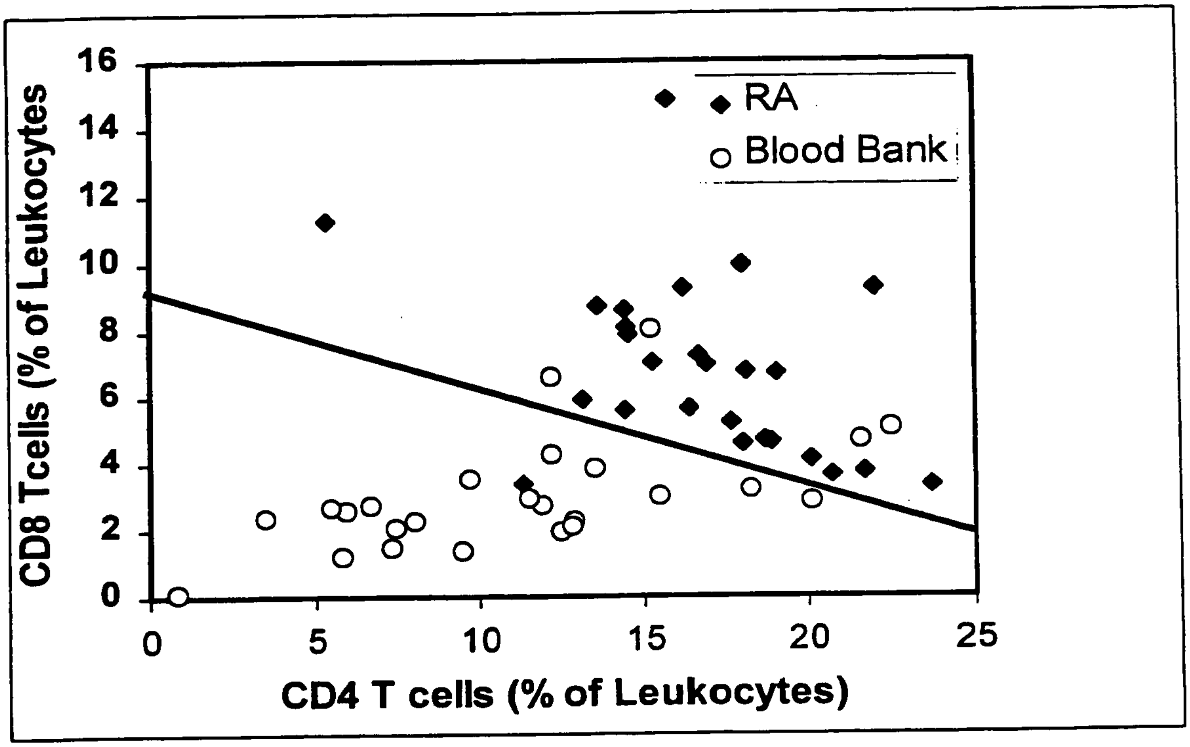


Figure 5B



RA
Blood Bank
CD8 Tcells (% of Leukocytes)
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
CD4 T cells (% of Leukocytes)

Figure 6

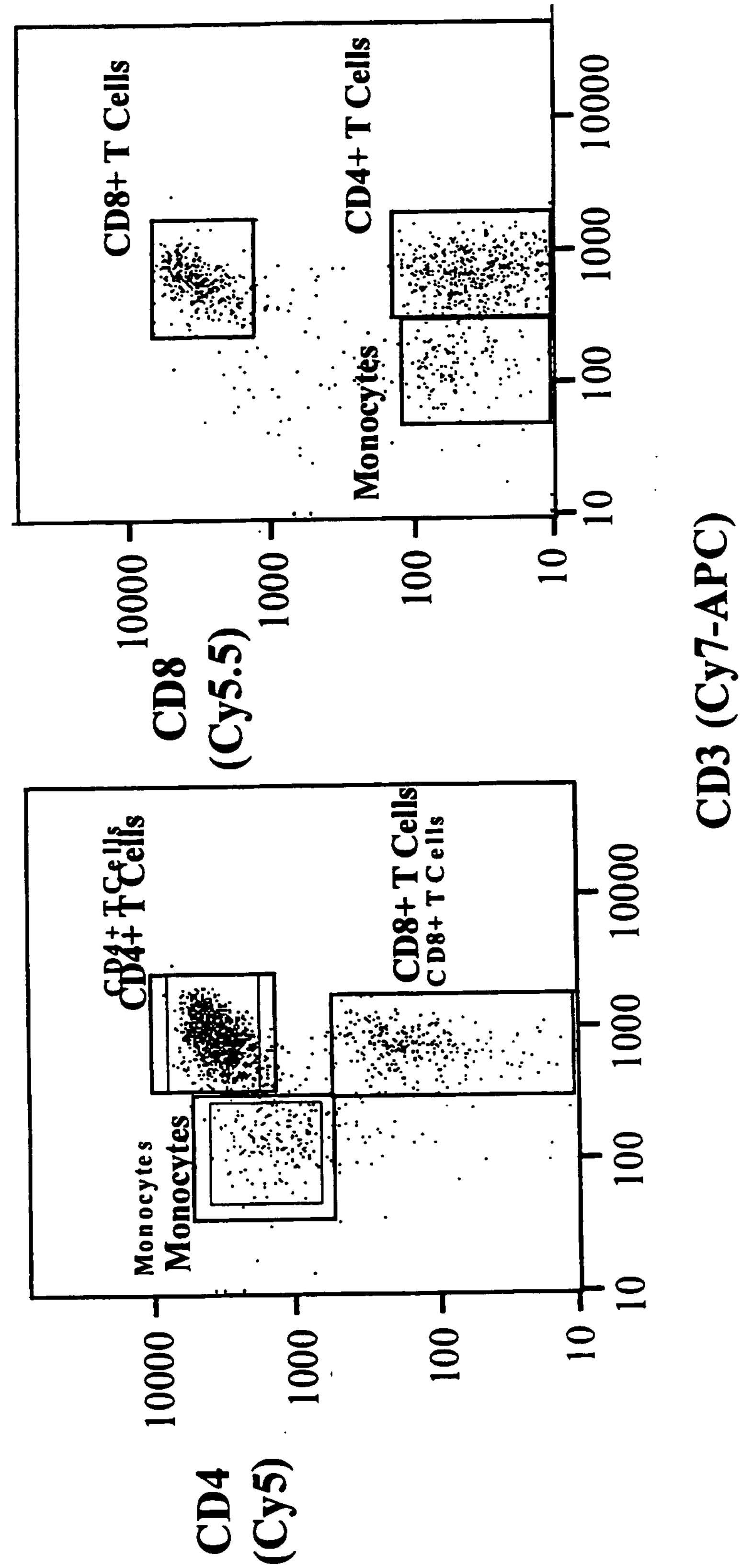


Figure 7B

Figure 7A

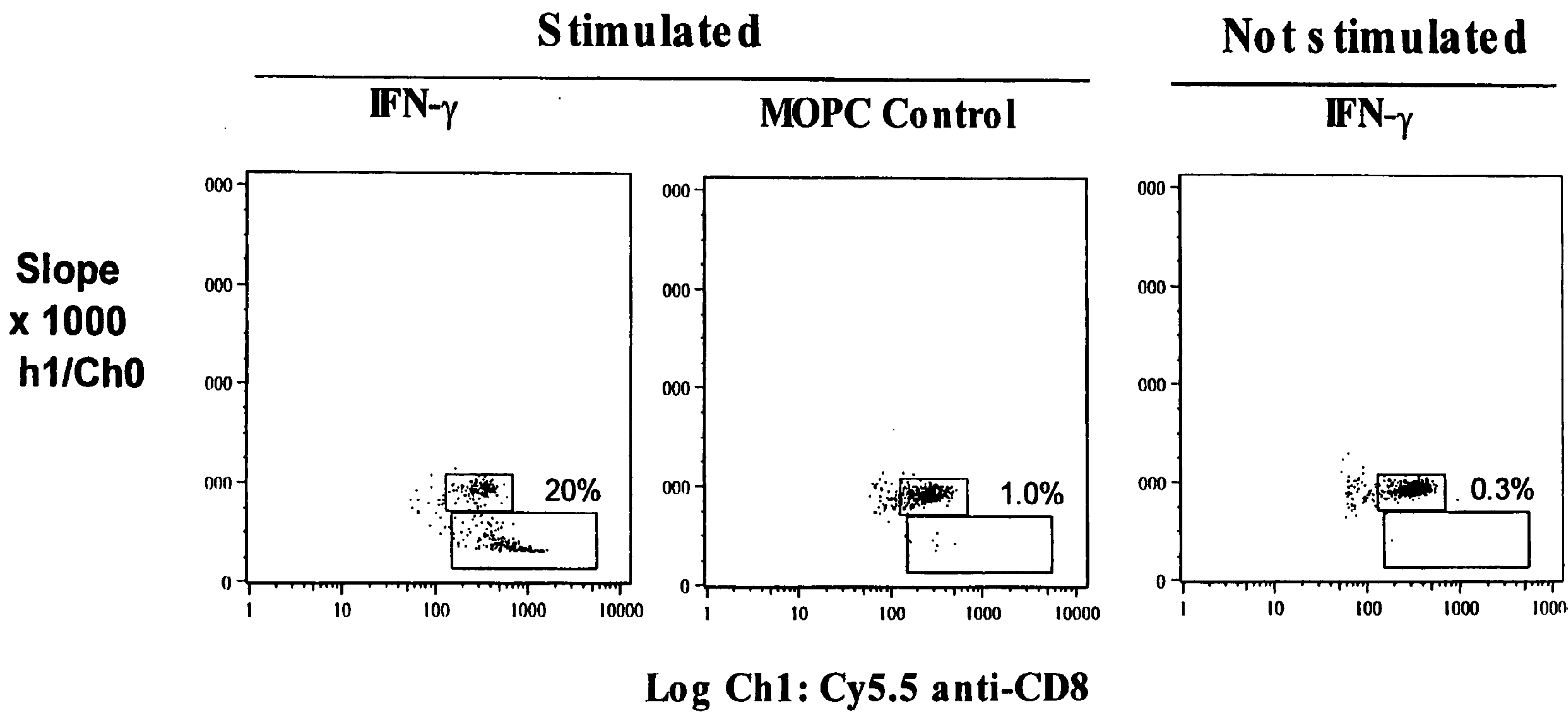


Figure 8A Figure 8B Figure 8C

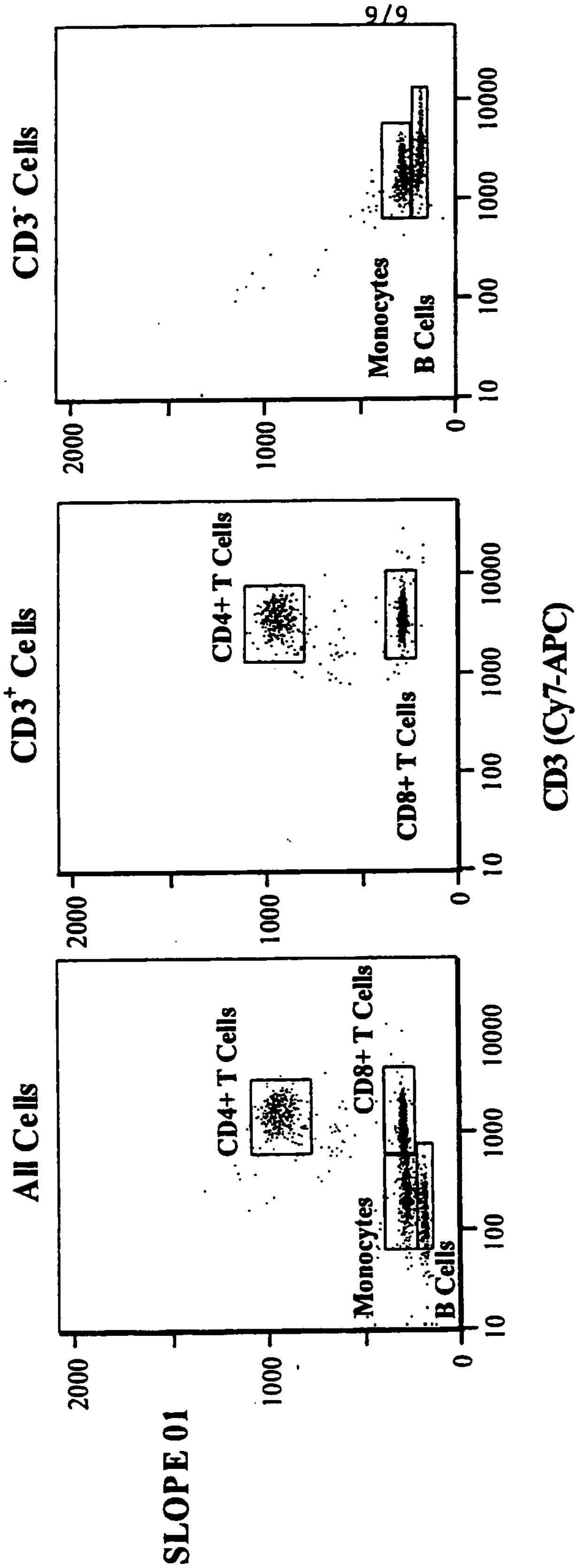


Figure 9A

Figure 9B

Figure 9C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/11296

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 17/00
US CL : 706/45

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 706/45

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Database Medline on STN, Department of Microbiology and Tumor Biochemistry Cancer Institute, (Madras, India), No. 90183328, VENKATANARAYANAN et al. 'Computerised algorithm of tumor-associated markers to monitor haematopoietic malignancy,' abstract, Computer Methods and Programs in Biomedicine, January 1990.	1-2, 4-5,7-19,22-57, and 60-71

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 19 June 2000 (19.06.2000)	Date of mailing of the international search report 25 JUL 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Young J. Kim Joyce Bridges Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/11296

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 f first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claim Nos.: 3,6,20-21,and 58-59
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by clai..s Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

Continuation of B. FIELDS SEARCHED Item 3: STN Commercial Database (Biosis, Medline, Embase, Embal, Scisearch, BiotechDS, Caplus)

Search Terms: Database, Dbase, LIMS, biological information, parameters, markers